DOCUMENT RESUME

ED 274 022	CS 505 394

AUTHOR	O'Brien, Nancy, Ed.
TITLE	Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, January 1-March 31, 1986.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Institutes of Health (DHHS), Bethesda, Md.; National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C.
REPORT NO	SR-85(1986)
PUB DATE	86
CONTRACT	NICHHD-NO1-HD-5-2910; ONR-N00014-83-K-0083
GRANT	NICHHD-HD-0109904; NIH-BRS-RR-05596; NINCDS-NS-13617; NINCDS-NS-13870; NINCDS-NS-18010; NSF-8520709; NSF-BNS-8111470
NOTE	273p.
AVAILABLE FROM	U.S. Department of Commerce, National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22151.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143) -- Viewpoints (120)

EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	*Communication Research; *Deafness; Morphophonemics; Phonology; Reading Research; *Speech Communication

ABSTRACT

The articles in this report explore the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical research applications. Titles of the papers and their authors are as follows: (1) "Phonological Awareness: The Role of the Reading Experience" (Virginia A. Mann); (2) "An Investigation of Speech Perception Abilities in Children Who Differ in Reading Skill" (Susan Brady, Erica Poggie, and Michele Merlo); (3) "Phonological and Morphological Analysis by Skilled Readers of Serbo-Croatian" (Laurie B. Feldman); (4) "Visual and Production Similarity of the Handshapes of the American Manual Alphabet" (John T. Richards and Vicki L. Hanson); (5) "Short-term Memory for Printed English Words by Congenitally Deaf Signers: Evidence of Sign-Based Coding Reconsidered" (Vicki L. Hanson and Edward H. Licktenstein); (6) "Morphophonology and Lexical Organization in Deaf Readers" (Vicki L. Hanson and Deborah Wilkenfeld); (7) "Perceptual Constraints and Phonological Change: A Study of Nasal Vowel Height" (Patrice Streeter Beddor, Rena Arens Krakow, and Louis M. Goldstein); (8) "The Thai Tonal Space" (Arthur S. Abramson); (9) "P-Centers Are Unaffected by Phonetic Categorization" (Andre Maurice Cooper, D. H. Whalen, and Carol Ann Fowler); (10) "Two Cheers for Direct Realism" (Michael Studdert-Kennedy); (11) "An Event Approach to the Study of Speech Perception from a Direct-Realist Perspective" (Carol A. Fowler); (12) "The Dynamical Perspective on Speech Production: Data and Theory" (J. A. S. Kelso, E. L. Saltzman, and B. Tuller); (13) "The Velotrace: a Device for Monitoring Velar Position" (Satoshi Horiguchi and Fredericka Bell-Berti); (14) "Towards an Articulatory Phonology" (Catherine P. Browman and Louis M. Goldstein); (15) "Representation of Voicing Contrasts Using Articulatory Gestures" (Louis Goldstein and Catherine P. Browman); and (16) "Mainstreaming Movement Science" (J. A. S. Kelso). (HTH)

ED274022

SR-85 (1986)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on the Nature of Speech, Instrumentation for its Investigation, and Practical Applications

1 January - 31 March 1986

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06511

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

(The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order number of previous Status Reports.)

CS505394

Ignatius G. Mattingly, Acting Editor-in-Chief
Nancy O'Brien, Editor

ACKNOWLEDGMENTS

The research reported here was made possible
in part by support from the following sources:

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Grant HD-01994

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Contract NO1-HD-5-2910

NATIONAL INSTITUTES OF HEALTH

Biomedical Research Support Grant RR-05596

NATIONAL SCIENCE FOUNDATION

Grant BNS-8111470
Grant BNS-8520709

NATIONAL INSTITUTE OF NEUROLOGICAL AND COMMUNICATIVE DISORDERS AND STROKE

Grant NS 13870
Grant NS 13617
Grant NS 18010

OFFICE OF NAVAL RESEARCH

Contract N00014-83-K-0083

HASKINS LABORATORIES PERSONNEL IN SPEECH RESEARCH

Investigators

- Arthur S. Abramson*
- Peter J. Alfonso*
- Thomas Baer
- Fredericka Bell-Berti*
- Catherine Best*
- Geoffrey Bingham†
- Gloria J. Borden*
- Susan Brady*
- Catherine P. Browman
- Franklin S. Cooper*
- Stephen Crain*
- Robert Crowder*
- Laurie B. Feldman*
- Anne Fowler†
- Carol A. Fowler*
- Louis Goldstein*
- Vicki L. Hanson
- Katherine S. Harris*
- Amelia I. Hudson[1]
- Leonard Katz*
- J. A. Scott Kelso*
- Andrea G. Levitt*
- Alvin M. Liberman*
- Isabelle Y. Liberman*
- Leigh Lisker*
- Virginia Mann*
- Ignatius G. Mattingly*
- Nancy S. McGarr*
- Richard S. McGowan
- Kevin G. Munhall
- Hiroshi Muta[2]
- Susan Nittrouer††
- Patrick W. Nye
- Lawrence J. Raphael*
- Bruno H. Repp
- Philip E. Rubin
- Elliot Saltzman
- Donald Shankweiler*
- Michael Studdert-Kennedy*
- Betty Tuller*
- Michael T. Turvey*
- Douglas H. Whalen

Technical/Support

- Philip Chagnon
- Alice Dadourian
- Michael D'Angelo
- Betty J. Delise
- Vincent Gulisano
- Donald Hailey
- Raymond C. Huey*
- Sabina D. Koroluk
- Yvonne Manning
- Bruce Martin
- Nancy O'Brien
- William P. Scully
- Richard S. Sharkany
- Edward R. Wiley

Students*

- Joy Armson
- Dragana Barac
- Eric Bateson
- Suzanne Boyce
- Teresa Clifford
- Andre Cooper
- Margaret Dunn
- Carole E. Gelfer
- Bruce Kay
- Noriko Kobayashi
- Rena A. Krakow
- Deborah Kuglitsch
- Hwei-Bing Lin
- Katrina Lukatela
- Harriet Magen
- Sharon Manuel
- Jerry McRoberts
- Lawrence D. Rosenblum
- Arlyne Russo
- Richard C. Schmidt
- John Scholz
- Robin Seider
- Suzanne Smith
- Katyanee Svastikula
- David Williams

---

*Part-time

[1] Visiting from Louisiana State University, Baton Rouge, LA

[2] Visiting from University of Tokyo, Japan

†NIH Research Fellow

††NRSA Training Fellow

CONTENTS

PHONOLOGICAL AWARENESS: THE ROLE OF READING EXPERIENCE
Virginia A. Mann 1-22

AN INVESTIGATION OF SPEECH PERCEPTION ABILITIES IN CHILDREN WHO DIFFER IN READING SKILL
Susan Brady, Erika Poggie, and Michele Merlo 23-37

PHONOLOGICAL AND MORPHOLOGICAL ANALYSIS BY SKILLED READERS OF SERBO-CROATIAN
Laurie B. Feldman 39-50

VISUAL AND PRODUCTION SIMILARITY OF THE HANDSHAPES OF THE AMERICAN MANUAL ALPHABET
John T. Richards and Vicki L. Hanson 51-64

SHORT-TERM MEMORY FOR PRINTED ENGLISH WORDS BY CONGENITALLY DEAF SIGNERS: EVIDENCE OF SIGN-BASED CODING RECONSIDERED
Vicki L. Hanson and Edward H. Lichtenstein 65-71

MORPHOPHONOLOGY AND LEXICAL ORGANIZATION IN DEAF READERS
Vicki L. Hanson and Deborah Wilkenfeld 73-84

PERCEPTUAL CONSTRAINTS AND PHONOLOGICAL CHANGE: A STUDY OF NASAL VOWEL HEIGHT
Patrice Streeter Beddor, Rena Arens Krakow, and Louis M. Goldstein 85-104

THE THAI TONAL SPACE
Arthur S. Abramson 105-114

P-CENTERS ARE UNAFFECTED BY PHONETIC CATEGORIZATION
André Maurice Cooper, D. H. Whalen, and Carol Ann Fowler 115-131

TWO CHEERS FOR DIRECT REALISM
Michael Studdert-Kennedy 133-138

AN EVENT APPROACH TO THE STUDY OF SPEECH PERCEPTION FROM A DIRECT-REALIST PERSPECTIVE
Carol A. Fowler 139-169

THE DYNAMICAL PERSPECTIVE ON SPEECH PRODUCTION: DATA AND THEORY
J. A. S. Kelso, E. L. Saltzman, and B. Tuller 171-205

THE VELOTRACE: A DEVICE FOR MONITORING VELAR POSITION
Satoshi Horiguchi and Fredericka Bell-Berti 207-218

TOWARDS AN ARTICULATORY PHONOLOGY
Catherine P. Browman and Louis M. Goldstein 219-250

REPRESENTATION OF VOICING CONTRASTS USING ARTICULATORY GESTURES
Louis Goldstein and Catherine P. Browman 251-254

MAINSTREAMING MOVEMENT SCIENCE
J. A. S. Kelso 255-262

PUBLICATIONS 265-266

APPENDIX: DTIC and ERIC numbers (SR-21/22 - SR-84) 267-268

7

Status Report on Speech Research

Haskins Laboratories

PHONOLOGICAL AWARENESS: THE ROLE OF READING EXPERIENCE*

Virginia A. Mann†

Abstract. A cross-cultural study of Japanese and American children has examined the development of awareness about syllables and phonemes. Using counting tests and deletion tests, Experiments I and III reveal that in contrast to first graders in America, most of whom tend to be aware of both syllables and phonemes, almost all first graders in Japan are aware of mora (phonological units roughly equivalent to syllables), but relatively few are aware of phonemes. This difference in phonological awareness may be attributed to the fact that Japanese first graders learn to read a syllabary, whereas American first graders learn to read an alphabet. For most children at this age, awareness of phonemes may require experience with alphabetic transcription, whereas awareness of syllables may be facilitated by experience with a syllabary, but be less dependent upon it. To clarify further the role of knowledge of an alphabet on children's awareness of phonemes, Experiments II and IV administered the same counting and deletion tests to Japanese children in the later elementary grades. Here the data reveal that many Japanese children become aware of phonemes by age ten whether or not they have received instruction in alphabetic transcription. Discussion of these results focuses on some of the other factors that may promote phonological awareness.

Introduction

The primary language activities of listening and speaking do not require an explicit awareness of the internal phonological structure of words any more than they require an explicit awareness of the rules of syntax. Yet a "metalinguistic" awareness that words comprise syllables and phonemes is precisely what is needed when language users turn from the primary language activities of speaking and listening to the secondary language activities of reading, versification, and word games (Liberman, 1971; Mattingly, 1972,

---

*Cognition, in press.

†Also Bryn Mawr College.

Acknowledgment. The research reported in this paper was completed while the author was a Fulbright Fellow and was partially funded by NICHD Grant HD21182-01 and by NICHD Grant HD01994 to Haskins Laboratories, Inc. This study could not have been completed without the help of Dr. Seishi Hibi, who served as research assistant, and without the very gracious compliance of Ms. Shizuko Fukuda and the children and teachers of the primary school attached to Ochanomizu University. I am also indebted to Dr. M. Sawashima, Dr. Isabelle Liberman, Dr. T. Ueno, and Dr. S. Sasanuma for their advice during many stages of this project.

1984). While all members of a given community become speakers and hearers, not all become readers, nor do they all play word games or appreciate verse. This difference raises the possibility that the development of phonological awareness might require some special cultivating experience above and beyond that which supports primary language acquisition.

Several different research groups have reported that adults who cannot read an alphabetic orthography are unable to manipulate phonemes (Byrne & Ledez, 1986; Liberman, Rubin, Duquès, & Carlisle, 1985; Morais, Cary, Alegria, & Bertelson, 1979; Read, Zhang, Nie, & Ding, 1984), raising the possibility that knowledge of the alphabet is essential to awareness of phonemes. In further pursuit of the factors that give rise to phonological awareness, the present study has explored the awareness of syllables and phonemes among Japanese children and American children. This particular cross-linguistic comparison is prompted by certain differences between the English and Japanese orthographies, and by certain differences in the word games and versification devices that are available to children in the two language communities.

Children in America learn to read the English orthography, an alphabet that represents spoken language at the level of the phoneme. Many of them also play phoneme-based word games such as "pig-Latin" and "Geography," and learn to employ versification devices such as alliteration that involve manipulations of phonemes, as well as word games and versification devices that exploit meter and thus operate on syllable-sized units. In contrast, virtually all of the secondary language activities that are available to Japanese children manipulate mora--phonological units that are roughly equivalent to syllables--if they manipulate phonological structure at all. Japanese children learn to read an orthography that comprises two types of transcription: Kanji, a morphology-based system, and Kana, a phonology-based system. Kanji is derived from the Chinese logography and represents the roots of words without regard to grammatical inflections, whereas Kana is of native origin and comprises two syllabaries, Hiragana and Katakana, which can represent the root and inflection of any word in terms of their constituent mora. Typically, the two orthographies function together, with Kanji representing most word roots and Kana representing all word inflections and the roots of those words that lack Kanji characters. As for other secondary language activities, Japanese word games such as "Shiritori" (a mora-based equivalent of "Geography") and versification devices such as Haiku manipulate mora.

In short, Japanese secondary language activities do not manipulate language at the level of the phoneme, whereas several English secondary language activities are phoneme-based, most notably the alphabetic orthography. Both Japanese and English afford versification devices and word games that manipulate syllable- sized units, but the Japanese orthography is unique in its inclusion of a syllabary. Given these similarities and differences between the orthographies and other secondary language activities in English and Japanese, it may be reasoned that, if experience with secondary language activities plays a specific role in the development of awareness about syllables and phonemes, Japanese children should be aware of mora (syllables), whereas American children should be aware of both phonemes and syllables. Should the experience of learning to read a given type of orthography play a particularly critical factor, Japanese children should be more aware of syllables than their American counterparts, who should be more aware of phonemes. It seems unlikely that the possession of primary language skills is sufficient to make Japanese and American children equivalent in awareness of phonemes, given findings that alphabet-illiterate adults are not aware of phonemes. However, it remains possible that children in the two countries will be equivalent in phonological awareness should reading experience or some other form of secondary language experience that draws the child's attention to the phonological structure of language promote the awareness of both syllables and phonemes.

The possibility that reading experience plays a particularly important role in the development of phonological awareness arises from the many studies that reveal an association between phonological awareness and success in learning to read an alphabetic orthography. These reveal that performance on tasks that require manipulations of phonological structure not only distinguishes good and poor readers in the early elementary grades (see, for example, Alegria, Pignot, & Morais, 1982; Fox & Routh, 1976; Katz, 1982; Liberman, 1973; Rosner & Simon, 1973) but also correlates with children's scores on standard reading tests (see, for example, Calfee, Lindamood, & Lindamood, 1973; Fox & Routh, 1976; Perfetti, 1985; Stanovich, Cunningham, & Freeman, 1984b; Treiman & Baron, 1983).

In many studies of reading ability and phonological awareness, the question of cause and effect has been broached, but never completely resolved. One of the earliest studies revealed that American children's awareness of phonological structure markedly improves at just that age when they are beginning to read (Liberman, Shankweiler, Fischer, & Carter, 1974): Among a sample of four, five, and six-year-olds, none of the youngest children could identify the number of phonemes in a spoken word, while half could identify the number of syllables; of the five-year-olds, 17 percent could count phonemes while, again, half could count syllables. Most dramatically, 70 percent of the six-year-olds could count phonemes and 90 percent could count syllables. Did the older children become aware of syllables and phonemes because they were learning to read, was the opposite true, or both?

Certain evidence suggests that phonological awareness can precede reading ability or develop independently. First of all, various measures of phoneme awareness and syllable awareness are capable of presaging the success with which preliterate kindergarten children will learn to read the alphabet in the first grade (see, for example, Bradley & Bryant, 1983; Helfgott, 1976; Jusczyk, 1977; Liberman et al., 1974; Lundberg, Oloffson, & Wall, 1980; Mann, 1984; Mann & Liberman, 1984; Stanovich, Cunningham, & Cramer, 1984a). Second, there is evidence that explicit training in the ability to manipulate phonemes can facilitate preliterate children's ability to learn to read (Bradley & Bryant, 1985). Third, the awareness of syllables, in particular, does not appear to depend upon reading experience, as the majority of preliterate children can manipulate syllables by age six without having been instructed in the use of a syllabary or an alphabet (Amano, 1970; Liberman et al., 1974; Mann & Liberman, 1984), and the ability to manipulate syllables is not strongly influenced by the kind of reading instruction, "whole-word" or "phonics," that children receive in the first grade (Alegria et al., 1982).

Other evidence, however, has revealed that at least one component of phonological awareness--awareness of phonemes--may depend on knowledge of an alphabet. As noted previously, several different investigators have reported that the ability to manipulate phonemes is markedly deficient in adults who cannot read alphabetic transcription. Awareness of phonemes is deficient among semi-literate American adults (Liberman et al., 1985), reading-disabled Australian adults (Byrne & Ledez, 1986), illiterate Portugese adults (Morais et al., 1979), and Chinese adults who can read only the Chinese logographic orthography (Read et al., 1984). In addition, the type of reading instruction that children receive can influence the extent of their awareness: first-graders who have been taught to read the alphabet by a "phonics" approach tend to be more aware of phonemes than those who have learned by a "whole-word" method (Alegria et al., 1982).

Present evidence, then, suggests that the relationship between phonological awareness and reading ability is a two-way street (Perfetti, 1985), which may depend on the level of awareness being addressed. Awareness of syllables is not very dependent on reading experience and could be a natural cognitive achievement of sorts, whereas awareness of phonemes may depend upon the experience of learning to read the alphabet, in general, and on methods of instruction that draw attention to phonemic structure, in particular. As a test of this view, the present study examined the phoneme and syllable awareness of children in a Japanese elementary school, predicting that these children would be aware of syllables, but would not be aware of phonemes until that point in their education when they receive instruction in the use of alphabetic transcription.

The design of the study involves four experiments that focus on the awareness of syllables (mora) and phonemes among children at different ages. Two different experimental paradigms are employed as a control against any confounding effects of task-specific variables. One paradigm is the counting test developed by Liberman and her colleagues, a test used in several studies of phonological awareness among American children (see, for example, Liberman et al., 1974; Mann & Liberman, 1984). The other is a deletion task, much like that employed by Morais et al. (1979) and Read et al. (1984) in their studies of alphabet-illiterate adults.

Experiment I used the counting test paradigm to study Japanese first-graders who had recently mastered the Kana syllabaries. To clarify the impact of knowledge of a syllabary vs. an alphabet, the results are compared with those reported in Liberman et al.'s (1974) study of American first graders. The relation between reading and phonological awareness is also probed by an analysis of the relation between phoneme and syllable counting performance and the ability to read Hiragana, in which case a nonlinguistic counting test guards against the possibility that any correlations might reflect attention capacity, general intelligence, etc. To further clarify the role of knowledge of the alphabet, Experiment II extended use of the counting test paradigm to Japanese children in the third to sixth grades. In Japan, children routinely receive some instruction in alphabetic transcription (Romaji) at the end of the fourth grade. There also exist certain "re-entry" programs for fourth through sixth graders who have spent the first few years of their education abroad and who have learned to read an alphabetic orthography. Comparisons among the re-entering pupils and normal pupils at various grade levels clarifies the relative contribution of alphabetic knowledge vs. knowledge of Kana and Kanji.

Experiment III used the deletion test paradigm to replicate and extend the findings of Experiment I. Aside from the change in procedure, its major innovation was to employ nonsense words as stimuli, constructing them in a fashion to permit parallel testing of first graders in Japan and in America.

Analysis of the results concerns performance on each deletion test in relation to reading experience and reading ability. Finally, Experiment IV used the same paradigm in a partial replication of Experiment II, comparing Japanese fourth graders who had not received instruction in Romaji with sixth graders who had been taught about Romaji one and a half years prior to the test session.

Experiment I

Methods

Subjects

The subjects were 40 children attending the first grade of the primary school attached to Ochanomizu University, twenty girls and twenty boys chosen at random from the available population and serving with the permission of their parents and teachers. Mean age was 84.4 months at the time of testing, which was the beginning of the second trimester of the school year. As a measure of Hiragana reading ability, each child rapidly read aloud a list of thirty high-frequency nouns, adjectives, and verbs (Sasanuma, 1978), and the total reading time and the number of errors were recorded. Each child was also rated by his or her teacher as above-average, average, or below-average in Kana reading ability.

Materials

The experiment employed three sets of materials designed to measure the ability to count three types of items: mora, phonemes, and 30° angles (a nonlinguistic unit). All three sets were modeled after the materials of Liberman et al. (1974): Each contained four series of training items that offered the child an opportunity to deduce the nature of the unit being counted, followed by a sequence of test items. In the mora counting test and phoneme counting test, all training and test items were common Japanese words that had been judged by four informants (a linguist, a speech scientist, a teacher of Japanese, and a librarian) to be readily familiar to young children. In the angle counting test, the items were simple line drawings of abstract designs and common objects. A more complete description of each test follows.

Mora counting test. Mora are rhythmic units of the Japanese language that more-or-less correspond to syllables. Each mora is either an isolated vowel, a vowel preceded by a consonant, an isolated [n], or the first consonant in a geminate cluster. A basic difference between mora and English syllables is that mora cannot contain consonant clusters, in general, or consonants in final position. It is further the case that a single syllable of English may correspond to two mora of Japanese. This owes to the fact that, in a Japanese word such as *hon*, [n] can be a mora, whereas [n] cannot be a syllable of English, and to the fact that differences in vowel duration (one or two mora) and consonant closure duration (normal or an extra mora) distinguish minimal pairs of Japanese words but are not contrastive in English.

In the mora-counting test, each training series contained three words: two-, three- and four-mora in length. Within the first three series, the words formed a progressive sequence, as in *hito* (man), *hitotsu* (one), hitotsubu (a grain or drop), but the words of the fourth series bore no such relation to each other [i.e., ima (now), kitte (stamp), chiisai (small)]. To introduce some of the complexities of Japanese phonology, the third series included a devoiced vowel, and the fourth included a long vowel and a geminate consonant. To avoid biasing the child's decision as to whether the task was to count the mora in a word (a phonological strategy) or the number of Kana characters needed to spell the word (a spelling strategy), the training items included only those mora that are spelled with a single character. Thus it was left ambiguous whether the task was to count orthographic units, or phonological ones.

The test sequence consisted of 14 two-mora words, 14 three-mora words, and 14 four-mora words presented in a fixed random order. They represented common combinations of mora including the nasal mora, geminate vowels, geminate consonants, and devoiced vowels. There were four VV words, two CVV words, six CVCV words and two CVC words in the two-mora pool; two VCVV words, two VVCV words, two CVVCV words, three CVCVCV words, two CVCVC words, two CVCCV words, and one CVCVV word in the three-mora pool, and four VCVCVCV words, two VCCVCV words, one VCVCCV word, four CVCVCVCV words, two CVCVCVV words, and one CVCCVCV word in the four-mora pool. As a probe for whether children were counting mora or orthographic units, three of the test items included one of the Japanese mora spelled with two characters.

Phoneme counting test. The design was analagous to that for the mora-counting test, but items manipulated the number of phonemes instead of the number of mora. The four training series contained a variety of the possible two-, three-, and four-phoneme sequences of Japanese, including nasal mora, devoiced vowels, long vowels and geminate consonants. Each of the first three contained a progressive sequence of items [i.e., ho (sail), hon (book), hone (bone)], whereas the fourth did not (i.e., ta (field), kau (buy), shita (under)]. The test sequence contained 14 two-phoneme words, 14 three-phoneme words, and 14 four-phoneme words arranged into a fixed random order. They comprised a broad sample of the permissible phoneme sequences in Japanese, including nasal mora, geminate consonants and vowels and devoiced vowels, which avoided systematic relationships between the number of phonemes a word contained, and either the number of mora in that word, or the number of Kana needed to spell it. There were four VV words, eight CV words, and two VC words in the two-phoneme pool; two VVV words, four VCV words, four CVV words, and four CVC words in the three-phoneme pool, and six CVCV words, two CVVV words, two VCCV words, two VCVV words, and two VVCV words in the four-phoneme pool.

Angle counting test. The materials were simple black and white line drawings that appeared on three by five inch cards. From one to three 30° angles were embedded in each drawing and the task was to count the number of these angles. In keeping with the design of the phoneme- and mora-counting tests, there were four series of training trials; in the first three series, the items were a progressive set of simple geometric shapes, but in the fourth they were objects that bore no systematic relationship to each other. The test sequence comprised drawings of objects, seven with one angle, seven with two angles and seven with three angles, arranged in a fixed random sequence.

Procedure

Prior to testing, the children were divided into two groups of ten girls and ten boys each. One group received the mora counting test, the other received the phoneme-counting test, and both received the angle-counting test at the onset of the session and the reading test at the end. The procedure for all three counting tests was the same. The instructor (a native speaker of Japanese) took two small hammers and told the child that they would be playing a "counting game." He then demonstrated the first training series in progressive order by saying each word in a normal fashion (or displaying each card) and then tapping the number of mora, phonemes, or angles. Next, the demonstration was repeated, with the child copying the instructor (saying each word first), and then items in the series were presented in a fixed random order, and the child responded without benefit of demonstration. If an error was made, the item was repeated and presentation of another randomized series followed. Otherwise, training proceeded to the next series, until, on completion of the fourth training series, the test items were presented and the child was instructed to "count" each item without the benefit of response feedback.

Results and Discussion

In evaluating children's responses on the mora and phoneme counting materials, two different scores were computed: the number of correct responses (as in Mann & Liberman, 1984), and a pass/fail score in which the criterion for passing was six consecutive correct responses (as in Liberman et al., 1974). Both appear in Table I along with mean age and mean reading scores for children in each group. The children who counted mora were equivalent to those who counted phonemes in terms of mean age, measures of reading ability, and performance on the angle-counting test ($\underline{p}>.05$). However, whereas scores on the mora-counting test approached ceiling, scores on the phoneme-counting test were considerably lower, $\underline{t}(38)=20.20$, $\underline{p}<.0001$. In addition, all of the children had passed the mora counting test, whereas only 10% had passed the phoneme counting test. The percentage of Japanese children who passed each test can be compared with the percentage of American first graders who had passed comparable tests in Liberman et al's original study: 90% for syllable counting, and 70% for phoneme counting. Apparently, first-grade children who have been educated in the use of the alphabet tend to perform better on the phoneme counting test than those who have not. Moreover, while children who have been educated in a syllabary might do slightly better on the syllable counting test, any difference is less dramatic. At present, no strong conclusion can be reached about these differences and their implications: Different test materials were used in the two countries, and children were not told explicitly to focus on the spoken word as opposed to its orthographic representation. Both problems are surmounted in Experiment III, which employed 1) a common set of materials in the testing of Japanese and American first graders, and 2) instructions to manipulate the sound pattern of each item.

Performance on each test gave indications of the influence of knowledge of Kana. In the mora-counting test, children appeared to deduce that the task involved counting orthographic units rather than counting phonological units. The majority gave an extra "tap" to the three items that contained a mora spelled with two characters instead of one, as if they were counting the number of characters needed to spell the word, instead of the number of mora. Other, much less frequent, errors on this test involved words that contained geminate consonants or long vowels, both of which tended to be underestimated and were missed only by the poorest readers of the group.

Table I

The Ability of Japanese First Graders to Count Mora vs. Phonemes

	SUBJECT GROUP	
	MORA COUNTING	PHONEME COUNTING
Phonological Counting		
Mean No. Correct (Max.=42)	38.1	18.1
Percentage Passing	100.0	10.0
Angle Counting		
Mean No. Correct (Max.=21)	11.9	11.8
Kana Reading Ability		
Mean speed (in sec.)	61.1	60.7
Mean errors (Max.=30)	1.6	1.8
Mean teacher rating (Good=1,avg.=2, poor=3)	1.9	2.0
Mean age (in months)	83.7	84.1

Analagous adherence to a "spelling strategy" can be found in children's responses to the phoneme-counting materials. During a post-hoc interview, some of the children reported that they had tapped the number of Kana characters needed to spell a given a word, and then added one to arrive at the correct response. Use of a "kana plus one" strategy could not allow children to reach the criteria of six consecutive correct responses, but it certainly inflated the number of correct responses. Items (N=25) for which the "Kana-plus-one" strategy yielded the appropriate response were correctly counted by an average of 55% of the children (which is significantly better than chance, $\underline{t}(24)=2.62$, $\underline{p}<.05$). In contrast, only an average of 38% had been correct on each item (N=17) for which that strategy yielded the incorrect response (which is significantly less than the percentage of children giving correct responses to the strategy-appropriate items, $\underline{t}(40)=5.4$, $\underline{p}<.001$, and not significantly better than chance, $\underline{p}>.05$).

A final concern of this experiment was the relation between performance on each counting test and the ability to read Kana. For the children who learned to count mora, the number of correct responses on the mora counting test was significantly related to teacher ratings, r(20)= .72, $p<.0001$, Hiragana reading speed, r(20)=.58, $p<.003$, and the number of errors, r(20)= -.47, $p<.02$, but not to age, sex, or performance on the angle counting test. This is consistent with Amano's (1970) report that mora counting ability is related to the acquisition of the first few Kana characters by pre-school children, and extends his finding to children in the first grade who possess considerably greater knowledge of the Kana syllabary. For the children who learned to count phonemes, the number of correct responses on the phoneme counting test was also significantly related to teacher ratings, r(20)=.56, $p<.005$, reading speed r(20)=.65, $p<.001$, and reading errors, r(29)= -.57, $p<.004$, but not to age, sex, or angle counting performance.

Thus it would appear that performance on the phoneme counting test is related to the ability to read Kana even though Kana does not represent phonemes in any direct way. As both phoneme and syllable counting performance are related to the ability to read Hiragana, just as they are related to the ability to read an alphabet, it is tempting to posit a general capacity for phonological awareness that is related to experience in reading any phonologically-based orthography. This capacity need not be part of general intelligence, given the results of some recent studies of American children (Mann & Liberman, 1984; Stanovich et al., 1984b), and the present finding that there is no significant correlation between measures of reading ability and performance on the angle counting test. It could be a general product of learning to read a phonological orthography rather than the cause of reading success, commensurate with children's reliance on Kana-based strategies. We will return to these issues in the final discussion.

The results of Experiment I are consistent with previous reports that awareness of phonemes depends on the experience of learning to read an alphabet, insofar as the majority of children could not pass the phoneme counting test. Nonetheless, two of the Japanese children did pass the test and our post-hoc interviews with them indicated that they had received no instruction in the alphabet either at home, school, or "juku" (i.e., afternoon training programs). Thus, while there may be some facilitating effects of learning a syllabary on awareness of both phonemes and syllables, some other factors may lead to individual variations. As a further test of the view that awareness of phonemes depends on the experience of learning to read an alphabet, we now turn to Experiment II, which focused on the phoneme counting ability of Japanese children in the third through sixth grades, comparing children at different grade levels in normal and "re-entering" classrooms.

Experiment II

Method

Subjects

The subjects were children attending the normal third- through sixth-grade classes and the special "re-entry" class at Ochanomizu University. The "normal class" subjects included 64 children in the third and fourth grades, and 32 children in the fifth and sixth grades. The "re-entry class" subjects included 13 fourth graders, 14 fifth graders, and 12 sixth graders, all of whom had learned to read either the English or German alphabet. Approximately equal numbers of boys and girls were included in each group and all served with parental permission. They were tested during the second trimester of school, so that children in the normal fourth-grade classes had not yet received training in the alphabet. Consultation with the teachers, the principal, and the children themselves confirmed that none of the subjects in normal classrooms had received instruction in the alphabet at school, home or "juku".

Materials and Procedure

The materials were the mora- and phoneme-counting materials employed in Experiment I, administered by the same instructor. For convenience, the procedure was adapted for group testing, in which case an entire class of children received the basic instructions and practice items with feedback, and learned to "count" each word by drawing slashes through the appropriate number of boxes in a five-box answer grid instead of by tapping the number of syllables/phonemes with a hammer. As in Experiment I, feedback was provided during training, but no feedback was provided during presentation of the test items. To insure the feasibility of group testing, the mora-counting materials were administered as a control measure to 32 of the third graders and 32 of the fourth graders. All of the remaining subjects received the phoneme-counting materials.

Results and Discussion

The data were scored in the manner of Experiment I, by computing both the number of correct responses and a pass/fail score. The results obtained from the mora-counting materials indicate the utility of the group testing procedure, as all of the third- and fourth-grade children had passed criterion with mean scores of 38.7 and 39.0, respectively. They also attest to the continuing power of the Kana orthography to mold the Japanese child's concept of language: As was the case in Experiment I, almost all of the children had made errors on the three test words in which the number of kana characters needed to spell the word surpasses the number of mora it contains.

Performance on the phoneme counting test is summarized in Table II, according to the age of the subjects, and whether they were in the normal or re-entry classes. On the basis of previous findings that alphabet-illiterate adults are not aware of phonemes, it might be expected that normal Japanese third and fourth graders would be no more aware of phonemes than the Japanese first graders studied in Experiment I, whereas the normal fifth and sixth graders and all of the re-entry students would be comparable to the American first graders studied by Liberman et al (1974). Yet, the data fail to uphold that prediction. First, for children in the normal classrooms, whose data appear in the upper portion of Table II, the only marked improvement in phoneme counting scores occurs between the third and fourth grades, prior to any instruction in the alphabetic principle. There is also no sharp spurt in the awareness of phonemes between fourth and fifth grades ($\underline{p}>.05$), such as would be expected if instruction in the alphabet were critical. Second, fourth graders in the reentry group performed at the same level as their peers in the normal classrooms ($\underline{p}>.05$), despite the fact that they alone had learned to read an alphabet. Third, and finally, the proportion of Japanese fourth graders who had passed criterion is comparable to that among the American children in Liberman et al.'s (1974) study, despite the fact that the Japanese children had not yet learned to read the Romaji alphabet.

Table II

Phoneme Counting Ability Among Japanese Children in the Third to Sixth Grades: Normal vs. Reentering Students

	Grade			
	Third	Fourth	Fifth	Sixth
Normal students				
Mean No. Correct (Max.=42)	21.5	30.3	31.2	31.5
Percentage Passing	56.2	73.5	81.3	75.0
Age (in months)	108.5	120.1	131.2	143.7
Reentering students				
Mean No. Correct (Max.=42)	---	27.2	28.6	27.7
Percentage Passing	---	60.0	60.0	80.0
Age(in months)	---	118.9	132.7	144.4

As in Experiment I, the importance of orthographic knowledge is illustrated by the pattern of errors, which suggests that at least some children were relying on the "Kana-plus-one" strategy of counting the number of characters needed to spell the word, and then adding one. Children at all ages tended to be most successful on items for which this strategy yielded the correct response: for strategy-appropriate items the average percent correct was 58%, 80%, 81%, and 82%, for third through sixth graders, respectively, whereas that for the strategy-inappropriate items was 42%, 56%, 64%, and 67%, respectively. Here, however, performance on both types of items surpassed the chance level of 33% correct ($\underline{p}<.05$), suggesting that appreciably many children at each age had been counting phonemes.

A popular organization of the Kana syllabary places the characters in a grid with the vowel mora in a different column to the far right of those containing characters for other mora. This organization had led us to anticipate that some of the subjects in Experiments I and II would use a strategy of giving the vowel mora one count and all other mora two counts. However, in post-hoc interviews of our subjects we found that none of them described such a strategy. Likewise, none of the children reported special treatment of the kana that can receive diacritics to mark the voicing of an initial stop consonant or fricative. Certainly it is possible that knowledge of Kana may have in some other way provoked children to reflect on the internal structure of words and thereby promoted phoneme awareness, but we were unable to determine why. Although children master Kana by the very early stages of first grade, the sharpest increase in phoneme counting performance occurs between third and fourth grade. Either increased experience of a very general sort or some maturational factors could be responsible.

In summary, although the findings of Experiment I suggest that both phoneme and syllable counting ability in the first grade might be facilitated by knowledge of an orthography that transcribes language at the level of that unit, the findings of Experiment II suggest that, analagous to the many American children who become aware of syllables by age six without having learned to read a syllabary, many Japanese children may become able to count phonemes by age nine or ten, despite a lack of formal instruction in the alphabet. Moreover, at that age, training in the use of an alphabet does not particularly enhance the ability to count phonemes. This finding stands in contrast to findings that most alphabet-illiterate adults appear to lack an awareness about phonemes.

One possible explanation of the performance differences between alphabet-illiterate adults and Japanese children is that they reflect task differences rather than differences in phonological awareness, per se. Japanese children might appear to be more aware of phonemes because the counting tasks employed in Experiments I and II were not explicit as to whether "sounds" or characters were to be counted, leading to reliance on a Kana-based strategy that inflated the number of correct responses. However, use of such a strategy could not account for changes in the percentage of children who passed the phoneme counting test, which raises the possibility that children passed the test because it provided a less conservative measure of phoneme awareness than the deletion tasks used in studies of adults. The results of at least one study are commensurate with this latter possibility. Performance on counting tasks and deletion tasks emerged as separate factors in a study of the relation between phonological awareness and the reading progress of semi-literate adults enrolled in a remedial reading class (Read & Ruyter, 1985). Another study, however, reveals that task-differences are not of critical importance to the relation between phonological awareness and the future reading success of kindergarten children in America (Stanovich et al., 1984a). However, as this latter study did not include counting tests, it remains a possibility that performance on counting tasks involves a more accessible level of phonological awareness than performance on deletion tests, hence the apparently greater awareness of phonemes on the part of Japanese children relative to alphabet-illiterate adults.

If the above explanation is correct, the present findings should not extend to use of a deletion test. On such a test, Japanese children should behave as poorly as alphabet-illiterate adults. With this prediction in mind, we turn to Experiments III and IV, which attempted to replicate Experiments I and II with deletion tasks analagous to those employed by Morais et al. (1979) and by Read et al. (1984). Two sets of nonsense-word materials were designed, one for phoneme deletion and one for mora deletion. Nonsense words had been among the most difficult items for the adult subjects and therefore offer a maximally conservative measure of children's performance; they also permit parallel testing of Japanese and American children.

Experiment III

Method

Subjects

The subjects were 40 Japanese first graders and 40 American first graders. There were equally as many girls as boys, all of whom served with parental and teacher permission. The Japanese children were drawn from an available population of children who had not participated in Experiment I. Mean age was 84.4 months at the time of testing, which was midway through the second trimester of the school year. The American children were comparable in age and SES, and were attending the Bolles Primary School in Jacksonville, FL. Mean age was 84.1 months at the time of testing, which was early in the second semester of the school year. Measures of children's reading ability were obtained by having the teachers rate each child as good, average, or poor in reading ability, and by giving each child a test of word decoding skill: the Hiragana reading test described in Experiment I for Japanese children, and the Word Identification and Word Attack Subtests of the Woodcock Reading Mastery Test (Woodcock, 1973) for American children.

Materials

As in Experiment I, two parallel sets of materials were designed, one for assessing syllable deletion ability and one for assessing phoneme deletion ability. The design of each was prompted by the methodology of Morais et al. (1979) and Read et al. (1984): Each set of materials assessed deletion of two different tokens of the segment of interest, with blocked sequences of training items followed by test items. To make the items suitable for use in English and Japanese, it was necessary that they contain only those Japanese mora that bear a one-to-one relationship to English syllables. Thus, all items contained consonants and vowels shared by the two languages, and none of them contained long vowels, syllabic [n], geminate consonants, diphthongs, consonant clusters, or syllable-final consonants. Each test item, and the item formed by removing its initial mora (or phoneme, as appropriate), was judged to be meaningless in Japanese (by the informants who judged the items of Experiment I) and in English (by comparable English-speaking informants).

Syllable materials. These materials assessed children's ability to remove an initial syllable (mora), [ta] or [u], from a three-syllable/three-mora nonsense word. Twenty items started with [ta] and twenty with [u]; the second and third syllable of each word varied freely. For the purpose of testing, the items were blocked with respect to initial syllable, and each block was subdivided into ten practice items and ten test items.

Phoneme materials. These materials assessed children's ability to remove an initial phoneme, [ʃ] or [k], from a four- or six-phoneme (i.e., two or three syllables/mora) nonsense word. Twenty items started with [ʃ] and twenty with [k]. The second phoneme of each word was always one of the five permissible vowels such that, across the items, each initial phoneme was followed by each vowel once in a four-phoneme word, and once in a six-phoneme word, with the remaining portion of each item varied freely. For the purpose of testing, the items were blocked with respect to initial phoneme, and each block was divided into ten practice items and ten test items (such that two- and three-syllable words were equally divided between practice and test items, as were the five vowels that could occur in the second-phoneme position).

Procedure

Children were tested individually by native speakers who used comparable instructions in the two languages. Within each country, half of the children received the syllable deletion test, half received the phoneme deletion test, and all received the reading test at the conclusion of the session. For each deletion test, presentation of practice and test trials was blocked with respect to initial segment (i.e., [ta] or [u], [ʃ] or [k]) with order counterbalanced across subjects. The instructor explained that the task involved repeating a word and then trying to say it without the first sound. He or she then proceeded to demonstrate the first five practice items: saying each word, repeating it, and then saying it without the first syllable or phoneme. Next, each of these was repeated and the child was requested to imitate the instructor by repeating the item and then saying it "without the first sound." Then the final five practice items were administered without benefit of demonstration, but with response feedback. Completion of the practice items was followed by the ten test items, which were administered without response feedback. Completion of the first block of trials was followed immediately by presentation of the second block of training and test items.

Results and Discussion

Attempts to remove the initial segment from each item were scored as correct or incorrect. The mean number of correct responses appear in Table III, separately for the American and Japanese children, according to the type and token of the segment being manipulated. When averaged across tasks and tokens, the scores of American children are slightly superior, $F(1,76)=7.31$, $p<.009$. With regard to the type of segment being deleted, children in both countries found the phoneme deletion task more difficult than the syllable (mora) deletion one, $F(1,76)=87.64$, $p<.0001$. However, the extent of difference between scores on the two tasks was greater for the Japanese children, $F(1,76)=13.01$, $p<.0006$. As compared to the American children, the Japanese children received higher scores on the syllable deletion task, $t(38)=2.73$, $p<.05$, but lower scores on the phoneme deletion task, $t(38)=-4.09$, $p<.01$. There were no significant effects of token differences, nor interactions between this manipulation and other factors.

Table III

Mora (Syllable) Elision Ability vs. Phoneme Ability:
A Comparison of First Graders in Japan and America

	Mora Elision		Phoneme Elision	
	[u]	[ta]	[ʃ]	[k]
Japanese Children				
Mora Group				
Mean No. Correct: (Max. = 10, Age = 83.8 mo.)	9.15	9.55		
Phoneme Group				
Mean No. Correct: (Max. = 10, Age = 85.1 mo.)			1.75	3.10
American Children				
Syllable Group				
Mean No. Correct: (Max. = 10, Age = 83.5 mo.)	8.90	8.80		
Phoneme Group				
Mean No. Correct (Max. = 10, Age = 84.8 mo.)			5.72	5.61

A further analysis considered the relations between phoneme and syllable deletion performance (summed across tokens) and reading ability in each country. As anticipated by the results of Experiment I, the mora deletion performance of the Japanese children was related to the speed, $r(20)=.69$, $p<.001$, and number of errors made on the Hiragana test, $r(20)=.72$, $p<.001$, and also to the teacher's ratings of reading ability, $r(20)=.54$, $p<.005$. Likewise, their phoneme deletion ability also proved to be related to speed, $r(20)=.37$, $p<.05$, and errors on the Hiragana test, $r(20)=.38$, $p<.05$, and to teacher ratings, $r(20)=.47$, $p<.02$. For the American children, phoneme deletion ability was related to the sum of raw scores on the Woodcock tests, $r(20)=.61$, $p<.005$, and to the teacher's ratings, $r(20)=.57$, $p<.008$, but syllable deletion ability was not related to either measure of reading ability. In neither language community was the age or sex of the first graders related to reading ability, mora deletion ability, or phoneme deletion ability ($p>.1$).

The relative superiority of the American children in the case of the phoneme deletion task corroborates previous indications that awareness about phonemes is faciliated by the learning of an alphabetic orthography. The analagous finding that Japanese children perform at a superior level on the syllable deletion task suggests that awareness about syllables may be likewise facilitated by learning to read a syllabary. Nonetheless, the finding that both Japanese and American children achieved higher levels of performance on the syllable deletion test than on the phoneme deletion test suggests that the ability to read a syllabary is less critical to awareness about syllables than the ability to read an alphabet is to awareness about phonemes. We now turn to Experiment IV, which attempted to replicate the findings of Experiment II regarding the contribution of orthographic knowledge to the phoneme deletion performance of Japanese children in normal fourth- and sixth-grade classrooms.

Experiment IV

Method

Subjects

The subjects were 20 fourth graders and 20 sixth graders attending the normal classes of the Ochanomizu Elementary School. Ten boys and ten girls from each grade were chosen at random from among the available pool of children who had not participated in Experiment II (i.e., those whose only experience with alphabetic instruction had occurred in school). All served with teacher and parental permission. Testing was conducted during the first trimester of the school year such that only the sixth graders had been educated in the use of an alphabetic orthography. Mean ages for each group were 117.1 and 142.5 months, respectively.

Materials and Procedure

The materials and procedure for Experiment IV were the phoneme deletion materials employed in Experiment III. The only innovation was that, at the completion of the test session, each subject was given two of the test items to which he or she had responded correctly and was asked to explain how the correct response had been derived. This provided a test of whether subjects had relied on either a Kana-based or a Romaji-based spelling strategy.

Results

The mean number of correct responses appears in Table IV, separated according to grade level and the phoneme token ([ʃ] or [k]) being manipulated. It can be seen that the performance of the sixth graders surpassed that of the fourth graders, F(1,38)= 18.49, $p<.0001$, consistent with the fact that only the sixth graders had learned to use alphabetic transcription. When the present results were compared with those obtained in Experiment III (and shown in Table III), it was found that both the Japanese fourth and sixth graders had surpassed the Japanese first graders in mean performance on the phoneme deletion task, $t(38)=4.08$, $p<.01$ for fourth graders, and $t(38)=4.53$, $p<.01$ for sixth graders. The Japanese fourth graders performed at the same level as the American first graders ($p>.1$), and the Japanese sixth graders had actually surpassed them, $t(38)=5.11$, $p<.01$.

Table IV

Phoneme Elision Performance Among Older Japanese Children

Grade in School	Phoneme Elision [ʃ]	[k]
Fourth Grade		
Mean No. Correct: (Max. = 10, Age = 117.1 mo.)	4.82	7.55
Sixth Grade		
Mean No. Correct: (Max. = 10, Age = 142.5 mo.)	8.33	10.00

To gain some appreciation of the Japanese children's knowledge of Romaji, we conducted an informal post-hoc interview with the five children who performed the best at each grade level. We found that none of the fourth-graders could read the nonsense test materials written in Romaji, whereas three of the sixth graders could do so. In contrast, although we had not asked the American children to try to read the test materials, they had been able to read an appreciable number of nonsense words on the Woodcock word-attack test. It may be remembered that the Japanese fourth graders had not received any instruction in Romaji, whereas the sixth graders had received approximately four weeks of instruction a full year and a half prior to the test session. The American first graders, on the other hand, had been receiving intensive phonics-based instruction in the use of the English alphabet for more than six months immediately prior to the test session.

A further analysis reveals an effect of token variations: Both fourth and sixth graders tended to give more correct responses to items that began with [k] than to those that began with [ʃ], F(1,38)=20.73, $p<.0001$. This may be explained by hypothesizing a "character-substitution" strategy based on the previously mentioned grid for representing the Kana syllabary as a matrix of rows and columns in which mora that share a vowel lie in the same row, and those that share a consonant lie in the same column. Within that matrix, the character for [a] is to the immediate right of that for [ka], [i] is to the immediate right of [ki], [u] to [ku], etc. Thus, children might be tempted to spell a word by replacing the first character with the character that lies to its immediate right on the matrix. Use of this strategy could cause [k] to be easier to delete than [ʃ] because characters containing [k] are immediately adjacent to those for isolated vowels, whereas most that contain [ʃ] are spelled with the character for [ʃi] with a subscripted character for [ya], [ye], [yu] or [yo] (according to the identity of the vowel). Moreover, they lie at the opposite end of the grid from the vowel characters, making it less obvious how to derive the character for the relevant vowel from that which represents the CV.

In this regard, we had actually asked children to explain how they had been able to arrive at a correct response. Of the fourth graders, seven were unable to describe their strategy at all, nine gave evidence of using the "character substitution strategy," and four subjects described a "phonological" strategy that more or less amounted to doubling the vowel of the first syllable in a word and then removing the initial consonant-vowel portion (i.e., making [ki-pi] into [ki-i-pi], and then deleting [ki] to yield [i-pi]. The children who reported the "phonological strategy" had achieved some of the best scores in their age group, and they tended to be equally accurate in their responses to items containing [k] and [ʃ]. As for the sixth graders, all of whom had been exposed to the alphabet, only four appeared to have employed the "character substitution strategy", and they achieved some of the lowest scores in their age group especially for items that began with [ʃ]. Fifteen of the remaining children reported some version of the "phonological strategy," and only a single child reported a strategy of using Romaji.

General Discussion

The present study asked whether Japanese children's awareness of syllables and phonemes differs from that of American children, as a consequence of their having learned to read a syllabary instead of an alphabet. The results clearly showed that Japanese children's approach to phonological counting and deletions tests is influenced by their reading experience. Knowledge of the Kana syllabary tended to confound performance on tasks that attempted to assess ability to manipulate phonological units, whether the tasks involved counting or deleting phonemes or syllables, and whether the instructions were ambiguous or explicit as to whether orthographic or sound units were being counted. Younger children in particular tended to manipulate the characters that spell a word rather than the phonological units that the characters transcribe. This tendency has previously been observed among American children (Ehri & Wilce, 1980) and has been one form of evidence that knowledge of an alphabet is responsible for phoneme awareness.

The results further reveal performance differences between first graders in Japan and America and illustrate that knowledge of a syllabary/logography as opposed to an alphabet can have a very specific effect on phonological awareness. Relative to first graders in Japan, first graders in America can more accurately count the number of phonemes in words and can more accurately remove the initial phonemes from nonsense words. Thus, the experience of learning to read an alphabet must facilitate children's awareness of phonemes at this age. The analagous finding that Japanese children can surpass American children in performance on tasks that call for syllable manipulation likewise reveals that experience with a syllabary can facilitate the awareness of syllables. However, children, in general, find syllable manipulation an easier task than phoneme manipulation, which suggests that the experience of learning to read a syllabary vs. an alphabet is not the sole determinant of phonological awareness.

What might the other determinants be? First of all, the development of phonological awareness may be a multi-faceted process that depends on the abstractness of the unit at issue. Syllables, as compared to phonemes, are isolable acoustic segments; they are more superficial, less encoded components of the speech signal. Thus it is reasonable that syllable awareness should be an easier, more natural achievement of such factors as cognitive maturation and primary language development, requiring less special cultivating experience than awareness of phonemes. The results of previous research favor this view (Liberman et al., 1974; Alegria et al., 1982; Read et al., 1984). While awareness of syllables may be a precursor of awareness of phonemes, it is not sufficient, given that some individuals can manipulate syllables but not phonemes. Previous research had suggested that the ability to manipulate phonemes depends on knowledge of an alphabet (Byrne & Ledez, 1986; Liberman et al., 1985; Morais et al., 1979; Read et al., 1984), but the present study suggests that other factors can also play a role.

The findings of Experiments II and IV emphasize the role of factors other than knowledge of the alphabet in the development of phoneme awareness, by revealing that, whereas most Japanese first graders could manipulate syllables but not phonemes, the majority of Japanese children were able to manipulate both syllables and phonemes by the fourth grade, whether or not they had been instructed in the use of an alphabet. Thus, with increasing age and educational experience, Japanese children may become more and more capable of manipulating phonemes whether or not they are alphabet-literate.

This finding stands in contrast to previous reports that adults who do not know how to read an alphabet are not aware of phonemes, and some explanation is required. We may disregard the possibility that the differences between Japanese children and the alphabet-illiterate adults are due to task differences rather than differences in phonological awareness, per se. A concern with this possibility prompted Experiments III and IV, which employed deletion tasks analagous to those used in previous studies of illiterate adults. The results obtained in these experiments are much the same as those obtained with the counting tasks employed in Experiments I and II. This accords with some other observations that the task-unique cognitive demands posed by different tests of phonological awareness do not appreciably confound conclusions about young children's phonological awareness and its role in reading acquisition (Stanovich et al., 1984a).

Perhaps a more reasonable interpretation is to accept the differences between the present findings and those obtained with alphabet-illiterate adults as differences in phonological awareness. We might then explore the possibility that other types of secondary language activity are responsible for the superior phonological awareness of the older Japanese children. One clear likelihood is that awareness of both syllables and phonemes is promoted by the experience of learning Kana, owing to the fact that it is a phonological orthography. This accords with the fact that many of the adults who proved deficient in phoneme awareness were functional illiterates (i.e., the American and Portugese adults). It would also accord with the correlations between Kana reading ability and both syllable and phoneme awareness, observed in Experiments I and III (although the correlation leaves causality ambiguous). It might seem inconsistent with certain findings (i.e. Experiment III and Mann, 1984) that syllable awareness fails to correlate with the ability to read the alphabet, but ceiling effects are a possible confounding factor. Other studies, however, have reported a correlation between syllable awareness and reading ability (see, for example, Mann & Liberman, 1984; Alegria et al., 1982).

A more serious problem with the view that knowledge of a phonological orthography promotes all aspects of phonological awareness concerns the lack of phoneme awareness among adult readers of the Chinese orthography (Read et al., 1984). As noted by Gelb (1963), Chinese, the most logographic of all the writing systems, is not a pure logographic system because from the earliest times certain characters have represented not words but phonological units. Many Chinese characters, the "phonetic compounds," are composed of a radical and a phonetic, each of which otherwise represents a word of the language. As noted by Leong (in press), the "fanqui" principal has been employed since 600 A.D. for decoding phonetic compounds, a strategy that calls for blending the first part (initial consonant) and the tone of the word represented by the phonetic with the final part (syllable rhyme) of the word represented by the radical. Thus a compound, e.g., composed of "t'u" and "l'iau," decodes as "t'iau." Several Chinese colleagues inform me that classical methods of education in the Chinese logography have explicitly called the reader's attention to the phonetic components. Moreover, although phonological changes have necessarily altered the relationship between phonetic compounds and the words they represent, one recent study reveals that the adult readers of Chinese make use of the phonetic insofar as they name low-frequency (but not high-frequency) characters that involve phonetic compounds faster than non-phonetic compound characters (Seidenberg, 1985). Likewise, adult readers of Chinese can use phonetic radicals productively (Fong, Horne, & Tzeng, 1986), to give consistent pronunciations for nonsense logographs composed of radicals and phonetics that do not co-occur. Given these findings, it is somewhat puzzling that exposure to phonetic compounds did not promote phonological awareness among Read et al.'s subjects, if exposure to any phonological orthgography facilitates phoneme awareness.

Putting aside the role of reading experience, it is possible that phoneme awareness is facilitated by some other secondary language experience that is available to Japanese children but not to the adults studied in Portugal and China. For Japanese children, the appropriate experience might involve learning to analyze or manipulate the phonological structure of spoken words while playing word games like "Shiritori" or while learning about Haiku. That the experience facilitating phonological awareness need not be limited to reading is evident from previous findings about the utility of explicit training in phonemic analysis (see Treiman & Baron, 1983, for example). Exposure to nursery rhymes and other poetry, for example, could help to explain why many American children are aware of syllables before they learn to read. But it would have to be argued that experience with such secondary language activities facilitates the development of all aspects of phonological awareness in a very general way, else how are we to explain the fact that Japanese children became able to manipulate phonemes despite a lack of experience with games and versification devices that directly manipulate phoneme-sized units? Even if it is postulated that any secondary language experience that manipulates phonological structure can give rise to awareness of both syllables and phonemes, there remains a problem insofar as meter and rhyme are exploited by both Chinese and Portugese verse, song lyrics, etc., and would probably have been available to the illiterate adults who nonetheless lacked phoneme awareness. A further problem arises from the fact that, in the present study, all of the children were familiar with the Kana syllabary and the same types of word games and versification devices, yet only a small minority of the first graders (10%) were able to count phonemes, whereas the majority of fourth graders could do so.

A similar argument can be made against the view that Japanese children knew about phonemes because they had seen signs, labels, etc. written in the Romaji alphabet. Any explanation that passive exposure to the Romaji alphabet is responsible for the phoneme awareness of Japanese children would have to account for the fact that all children are exposed to Romaji signs and logos, yet only those aged nine and older had profited from that exposure. It would also have to account for the fact that passive exposure to alphabetically-written material failed to promote phoneme awareness among the Portugese adults studied by Morais et al. (1979).

One final explanation of the differences between the present results and those obtained with alphabet-illiterate adults remains. The ability to manipulate both syllable and phoneme-sized units could be a natural concomitant of primary language development that is exploited by many secondary language activities such as reading, versification, and word games. But if this capacity is a natural concomitant of primary language, how can it be deficient in alphabet-illiterate adults? Perhaps the ability to manipulate phonemes tends to atrophy unless maintained by appropriate reading experience. It has often been speculated that children acquire their primary language with the aid of a language acquisition device that is not present in adults. That the capacity for manipulating phonemes could be part and parcel of a language acquisition device follows from a suggestion made by Mattingly (1984), in answer to the question of why readers might be able to gain access to the otherwise reflexive processes that support the processing of phonological structure in spoken language. He suggests that an ability to analyze the phonological structure of spoken words might serve to increase the language learner's stock of lexical entries, and this, together with some other evidence that children have a privileged ability to acquire new lexical entries (Carey, 1978), could lead to the speculation that children have a privileged ability to manipulate phonological structure that somehow facilitates their ability to engage in secondary language activities that involve manipulations of phonological units. The prevalence of this capacity in childhood could promote children's acquisition of phonological orthographies during their elementary school years and by postulating that this capacity in the absence of appropriate orthographic knowledge, one might explain the lack of phoneme awareness observed among alphabet-illiterate adults. However, this view is not without its problems, one being the fact that Japanese children could not do well on either the counting or elision tasks until relatively late in their childhood. Here, the cognitive demands of tests that are used to measure phoneme awareness and the confounding role of orthographic knowledge cannot be disregarded. Ongoing research with a broader battery of tests and a broader range of ages may further elucidate the basis of phonological awareness in the interplay between cognitive skills, primary language skills, and experience with secondary language activities such as reading.

References

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. Memory & Cognition, 10, 451-456.

Amano, K. (1970). Formation of the act of analyzing phonemic structure of words and its relation to learning Japanese syllabic characters. Japanese Journal of Education, 18, 12-25.

Bradley, L., & Bryant, P. E. (1983). Difficulties in auditory organization as a possible cause of reading backwards. Nature, 271, 746-747.

Bradley, L., & Bryant, P. E. (1985). Rhyme and reason in reading and spelling. Ann Arbor: University of Michigan Press.

Byrne, B., & Ledez, J. (1986). Phonological awareness in reading-disabled adults. Australian Journal of Psychology, 35, 185-197.

Calfee, R. C., Lindamood, P. O., & Lindamood, C. (1973). Acoustic-phonetic skills and reading--kindergarten through twelfth grade. Journal of Educational Psychology, 64, 293-298.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), Linguistic theory and psychological reality (pp. 264-293). Cambridge, MA: MIT Press.

Ehri, L. C., & Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. Applied Psycholinguistics, 1, 371-385.

Fong, S. P., Horne, R. Y, & Tzeng, O. J. (1986). Consistency effects with Chinese character and pseudocharacter naming tests. In S. H. Kao & R. Hoosain (Eds.), Linguistics, psychology and the Chinese language. Hong Kong: University of Hong Kong Press.

Fox, B., & Routh, D. K. (1976). Phonemic analysis and synthesis as word-attack skills. Journal of Educational Psychology, 69, 70-74.

Gelb, I. J. (1963). A study of writing. Chicago: University of Chicago Press.

Helfgott, J. (1976). Phonemic segmentation and blending skills of kindergarten children: Inplications for beginning reading acquisition. Contemporary Educational Psychology, 1, 157-169.

Jusczyk, P. (1977). Rhymes and reasons: Some aspects of children's appreciation of poetic form. Developmental Psychology, 13, 599-607.

Katz, R. (1982). Phonological deficiencies in children with reading deficiencies: Evidence from an object naming task. Doctoral Dissertation, Department of Psychology, University of Connecticut, Storrs, CT.

Leong, C. K. (in press). What does accessing a morphophonemic script tell us about reading and reading disorders in alphabetic scripts. Bulletin of the Orton Society.

Liberman, I. Y. (1971). Basic research in speech and the lateralization of language: Some implications for reading. Bulletin of the Orton Society, 21, 71-87.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. Bulletin of the Orton Society, 23, 65-77.

Liberman, I. Y., Rubin, H., Duquès, S., & Carlisle, J. (1985). Linguistic abilities and spelling proficiency in kindergarteners and adult poor spellers. In D. B. Gray & J. F. Kavanagh (Eds.), Biobehavioral measures of dyslexia. Parkton, MD: York Press.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten.

Mann, V. A. (1984). Longitudinal prediction and prevention of reading difficulty. Annals of Dyslexia, 34, 117-137.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. Journal of Learning Disabilities, 17, 592-598.

Mattingly, I. G. (1972). Reading, the linguistic process and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationship between speech and reading*. Cambridge, MA: MIT Press.

Mattingly, I. G. (1984). Reading, linguistic awareness and language acquisition. In J. Downing & R. Valtin (Eds.), *Linguistic awareness and learning to read* (pp. 9-25). New York: Springer-Verlag.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323-331.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Read, C., & Ruyter, L. (1985). Reading and spelling skills in adults of low literacy. *Remedial and Special Education*, *6*, 43-52.

Read, C., Zhang, Y., Nie, H., & Ding, B. (1984). *The ability to manipulate speech sounds depends on knowing alphabetic transcription*. Paper presented at the 23rd International Congress of Psychology, Acapulco, September, 1984.

Rosner, J., & Simon, D. P. (1971). The auditory analysis test: An initial report. *Journal of Learning Disabilities*, *4*, 384-392.

Sasanuma, S. (1978). *Token Test of differential diagnosis of aphasia*. Tokyo: Yaesu Rehabilitation Center.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1-30.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984a). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology*, *38*, 175-190.

Stanovich, K. E., Cunningham, A. E., & Freeman, D. J. (1984b). Intelligence, cognitive skills and early reading progress. *Reading Research Quarterly*, *14*, 278-303.

Treiman, R., & Baron, J. (1983). Phonemic-analysis training helps children benefit from spelling-sound rules. *Memory & Cognition*, *11*, 382-389.

Woodcock, R. W. (1973). *Woodcock Reading Mastery Test*. Circle Pines, MN: American Guidance Services.

AN INVESTIGATION OF SPEECH PERCEPTION ABILITIES IN CHILDREN WHO DIFFER IN READING SKILL

Susan Brady,† Erika Poggie,†† and Michele Merlo††

Abstract. Considerable evidence indicates that children who are poor readers have a phonetic coding deficit on linguistic short-term memory tasks. A previous study (Brady, Shankweiler, & Mann, 1983) had explored whether the initial perception of items might be the locus of the memory problem, and had demonstrated inferior speech perception abilities for poor readers with degraded stimuli. In the present study, the goal was to look more closely at perception under clear listening conditions. Third-grade good and poor readers were tested on a word repetition task with monosyllabic, multisyllabic, and pseudoword stimuli. Poor readers were significantly less accurate on the more demanding multisyllabic and pseudoword stimuli, though no group differences were obtained on speed of responding. The lack of reaction time differences between good and poor readers was corroborated on a control task in which verbal response time to nonspeech stimuli was measured. The reduced accuracy with clearly presented stimuli confirms the presence of subtle deficiencies in speech perception for children with reading difficulty and strengthens the hypothesis that poor readers' memory deficits may stem from less efficient encoding processes.

Evidence has been steadily mounting that the associates of early reading difficulty lie in the phonological domain. One of the central areas of research contributing to this evidence has involved studies of short-term memory (STM). Children with reading problems have repeatedly been observed to have deficient recall on STM tasks when compared with better reading peers. The role of phonological processes in this deficit has been implicated by several findings: First, the memory deficit for poor readers is observed only for stimuli that can be phonetically recoded such as letters, words, and pictures of nameable objects. When stimuli are presented for recall that are not easily given a phonetic code, good and poor readers perform comparably.

---

†Also University of Rhode Island.

††University of Rhode Island.

Acknowledgment. We wish to thank several colleagues for their helpful comments and suggestions: Anne Fowler, Vicki Hanson, Joe Rossi, Richard Schmidt, and Donald Shankweiler. We are also grateful to those in the Narragansett Elementary School, Narragansett, Rhode Island, for their kind cooperation: William Holland, Superintendent; David Hayes, Principal; Judy Aiello, Reading Coordinator; the third grade teachers (Sue Boland, Leslie Flynn, Hope Rawlings, Gloria Sandel, and Marguerite Strain); and the wonderful children who worked so diligently. The research and the preparation of the manuscript were supported in part by a grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994).

[HASKINS LABORATORIES: Status Report on Speech Research SR-85 (1986)]

This contrasting result has been obtained with tasks employing photographs of strangers, nonsense doodle drawings, symbols from an unfamiliar writing system, and with auditorily presented tones (Holmes & McKeever, 1979; Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982; Vellutino, Pruzek, Steger, & Meshoulam, 1973). Thus the limits in STM for children with reading difficulty are specific to tasks requiring phonetic coding.

Second, when the STM tasks consist of linguistic material, manipulations of phonetic dimensions of the stimuli generally affect the performance of young good readers more than that of young poor readers (Liberman & Shankweiler, 1979; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). With strings of phonetically distinct (nonrhyming) stimuli, good readers show the usual superior recall for verbal material. When the phonetic confusability is increased by presenting rhyming items, the performance of good readers is impaired much more than the recall of poor readers. It has been reasoned that this pattern, also observed in adults, stems from the skilled readers being better able to form a sufficient phonetic code for temporary storage of information. Stimuli that minimize the phonetic contrasts between items in STM, such as lists of rhyming words, thus tend to have a greater effect on the recall of the good readers. Therefore, differential sensitivity to phonetic similarity by reading groups has been seen as a consequence of differing levels of skill in the use of a phonetic code.

Subsequent studies have indicated that poor readers employ a phonetic code, but do so less accurately than good readers. Examining the nature of errors on verbal STM tasks, both reading groups produce phonetically-based mistakes such as transpositions of phonetic elements. However, the incidence of these errors is more frequent for the children with reading difficulty (Brady, Mann, & Schmidt, 1985; Brady, Shankweiler, & Mann, 1983). Additional research indicates that poor readers are not worse on all components of language processing:[1] when other linguistic variables in STM tasks are experimentally varied, such as syntactic and semantic parameters, reading groups are equally affected (Mann, Liberman, & Shankweiler, 1980). Thus the memory deficits of poor readers are uniquely associated with phonetic requirements in STM, not with other aspects of language processing.

An important insight about the extent of the phonetic coding problem arises from the observation that reading groups differ in STM recall whether the lists are presented visually or auditorily (Brady et al., 1983; Brady et al., 1985; Mann et al., 1980; Shankweiler et al., 1979). This finding suggests that poor readers experience a general difficulty in the use of a phonetic code, rather than an impairment specific to the encoding of visual information.

To summarize, poor readers demonstrate short-term memory deficits only for stimuli that are phonetically recodable. These children show reduced sensitivity to rhyme and greater frequency of phonetic errors of transposition, providing further support that the deficit is related to phonetic skills. Lastly, these results are independent of the modality of presentation, pointing to a general phonetic processing deficit in STM.

The current evidence is consistent with the view that the short-term memory deficit of poor readers stems from deficiencies in the use of a phonetic code. In exploring the phonetic basis of the memory problem, we have been conducting experiments to determine whether the problem arises in perception with the encoding of stimuli. If so, poor readers can be expected to do less well on perception as well as on recall tasks than good readers. This finding was obtained in a previous study (Brady et al., 1983) in which third-grade poor readers performed less accurately than good readers on a speech perception task requiring identification of words presented in noise. In contrast, the reading groups did not differ in performance on a nonspeech control task with environmental sounds.

At the present we are working with the hypothesis that the difficulties of poor readers in speech perception and verbal STM tasks arise from a common source: the creation and maintenance of phonetic representations. From this approach, the efficiency with which the input is encoded will have consequences both in perception and in memory. Rabbitt (1968) carried out experiments with adults that supported this hypothesis. When digits were degraded slightly by the addition of noise, memory was observed to suffer, even though identification of the digits in isolation was still accurate. Rabbitt proposed that limited processing capacity was the basis for the reduction in memory span. That is, as increased resources were required for identification of the digits in noise, relatively less processing capacity was available for retaining the items in memory.

Similar explanations have been offered for the commonly observed developmental increases in STM (Chi, 1976; Dempster, 1981), and the individual differences in memory span for adults (Baddeley, Thompson, & Buchanan, 1975; Hoosian, 1982). Hulme, Thomson, Muir, and Lawrence (1984) report that although younger children recall less, the same linear function relates speaking rate to short-term memory for subjects ranging in age from four years old to adulthood. They suggest that speech rate can be seen as a measure of rehearsal speed, so that increases in speech rate, rather than in memory span per se, account for the observed gains in STM during development. Case, Kurland, and Goldberg (1982) likewise found that speed of word repetition correlated with memory span scores for children three to six years of age. These authors propose the slightly different explanation that basic operations in perception and memory become more efficient with experience, requiring less processing space, and that as a consequence more functional space exists for storage. In an interesting test of this, Case et al. equated six year olds and adults on speed of word repetition by manipulating word familiarity, and correspondingly found that the word spans for these two age groups were no longer different.

Given that the efficiency of phonetic processes appears to be related to normal developmental increases in memory span, it is of added importance to evaluate whether the STM differences associated with reading ability might arise from the efficiency of phonetic encoding. Since poor readers in the Brady et al. (1983) study made more errors repeating the speech-in-noise, it appears their perceptual skills are less well developed than those of children who are good readers. Therefore, it might also be the case that under clear listening circumstances the poor reader's encoding, though adequate, is less efficient (i.e., may require more processing resources) and may limit performance on recall tasks.

To test this line of reasoning, we wanted to investigate whether differences in efficiency in perception are present under clear listening conditions, since this is the way stimuli are presented for most STM experiments. In the Brady et al. (1983) study, no perceptual differences between reading groups had been observed on accuracy scores for a noise-free task with monosyllabic words. However, since both good and poor readers had been at ceiling performance levels, it may not have been a sufficiently sensitive procedure to assess group differences in perceiving clear stimuli.

If reading group differences in perceptual processing efficiency are present for clear listening, we speculated that this might take one or two forms: 1) poor readers might be slower at identifying or producing a phonetic utterance; 2) the quality of the phonetic representation might be less fully accurate for the poor readers. Under clear listening conditions with no time constraints and with relatively easy phonetic stimuli, poor readers could conceivably perform well with either or both of these processing limitations.

With these questions in mind we examined third-grade good and poor readers on a speech repetition task. Our aim was to look more closely at whether reading group differences in perception are evident when stimuli are presented clearly. The responses were scored for accuracy, and reaction time (RT) measures were collected to assess processing speed. Three kinds of stimuli were presented: monosyllabic words, multisyllabic words, and pseudowords. In this way the phonological demands of the task were varied in case monosyllabic words (previously tested) were not sufficiently difficult to process to reveal potential group differences. Therefore the length of the stimuli was increased in the multisyllabic condition and the familiarity was decreased in the pseudoword condition. Both of these are known to increase processing demands in adults and we expected this also to be true for children. We hypothesized that the reduced accuracy of poor readers on speech-in-noise (Brady et al., 1983) had reflected on-going differences in perceptual skills that are only apparent on somewhat demanding tasks. Consequently we predicted that poor readers would be less accurate than good readers on the more difficult multisyllabic and pseudoword stimuli, but that both groups would do well on the monosyllabic items.

The reaction time measure allows us to address the speed of processing issue raised in the developmental literature (e.g., Hulme et al., 1984). If group differences in RT were evident, we predicted that the good readers would be faster, indicating more rapid phonetic processing capabilities and possibly reflecting a developmental advantage.

Anticipating that differences in reaction time might be present for good and poor readers, a control task was included so it would be possible to focus on what aspect of the repetition task was implicated. In this task, subjects were presented with nonspeech tones to which they were to respond rapidly with a specified word. If potential group RT differences in the word repetition task were related to articulation speed, reading group differences should be maintained on the control task. If instead they stemmed from identification processes for a phonetic input, the tone stimuli should not generate group RT differences.

Methods

Subjects. The subjects were third-grade children from a suburban school district in southern Rhode Island. The school reading coordinator targeted the children she thought would qualify as good or poor readers. These children were then administered the Word Attack and Word Recognition subtests of the Woodcock Reading Mastery Tests, Form A (Woodcock, 1973), and a test of receptive vocabulary, the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn, 1981). In addition, the children were screened for hearing loss. Using a standard audiometer, each child's right and left ears were tested with tones at 500 Hz (25dB), 1000 Hz (20dB), 2000 Hz (20 dB), 4000 Hz (20 dB) and 8000 Hz (20 dB).

Children were selected as subjects if they met the following criteria: (1) To ensure appropriate classification as a good or poor reader, an individual was included only if the two scores on the Woodcock subtests were consistent (i.e., if both scores indicated a comparable level of reading ability.) (2) In order to limit the range of vocabulary skills, participation was restricted to those with PPVT-R IQ scores between 90 and 125. (3) Because of the auditory requirements of the experimental tasks, only children who passed the hearing screening were eligible. In accord with routine procedures, an individual passed the screening if no more than a single frequency on each ear was undetected. (4) Given the evidence that the speech perception skills of children continue to progress during elementary school years (Finkenbinder, 1973; Goldman, Fristoe, & Woodcock, 1970; Schwartz & Goldman, 1974; Thompson, 1963), selection of subjects was limited to those whose ages fell within a one year span (101-113 mos.).

Thirty children (15 good readers and 15 poor readers) met the requirements for inclusion in the study. The characteristics of the two reading groups are summarized in Table 1. The Woodcock test scores were non-overlapping for the good and poor reading groups. The 15 children who were designated good readers were clearly beyond third grade reading mastery, with a mean reading grade level of 7.8. The 15 children who were labeled poor readers had an average lag of nine months below their expected level ($\bar{x}=3.1$). Neither the ages, $F(1,28)=.26$, $p = .61$, nor the PPVT-R IQ scores, $F(1,28) = 1.23$, $p = 2.8$, of the good and poor readers were significantly different.

Table 1

Means for Third Grade Children Grouped According to Reading Achievement

Group	N	Age	IQ[a]	Reading Grade[b]
Good	15	8 yr. 9 mo.	108.1	7.8
Poor	15	8 yr. 10 mo.	104.5	3.1

[a] Peabody Picture Vocabulary Test

[b] From the average of the reading grade scores obtained on the Word Attack and Word Recognition subtests of the Woodcock Reading Mastery Tests, Form A.

Stimuli. Three sets of stimuli were used: (1) a set of 48 monosyllabic words; (2) a set of 24 monosyllabic pseudowords, and (3) a set of 24 multisyllabic words. In addition, a 24 item control task was employed.

Monosyllabic words. The monosyllabic word list (MONO) was the same as that used in a previous study (Brady et al., 1983). The words were chosen to control for syllable pattern, phonetic composition, and word frequency. There were 12 words for each of four syllabic patterns: CVC (consonant-vowel-consonant), CCVC, CCVCC, and CVCC. In addition, the words were chosen to provide a systematic phonetic set. Twenty words began with stop consonants (/b/, /d/, /g/, /p/, /t/, /k/), twenty words began with fricatives, or affricates (/tʃ/, /s/, /f/, /ʃ/, /dz/, /v/), and four began with liquids (/r/, /l/). The same distribution of phonemes occurred in word final position.

For each syllable and phoneme pattern, half of the words included were reported to have a high frequency of occurrence in children's literature and half to have a low frequency (Carroll, Davies, & Richman, 1971). The words used are presented in Table 2.

Table 2

Monosyllabic Stimuli

Words		Pseudowords
High Frequency	Low Frequency	
door	bale	dar
team	din	tem
road	lobe	rud
knife	mash	nauf
chief	chef	chife
job	fig	jeeb
grain	tram	grun
breath	grouse	brath
crowd	crag	crad
sleep	slag	slape
scale	spire	skell
speech	skiff	spoach
front	flint	frant
plant	clamp	plint
friend	frond	freend
clouds	glades	cleeds
blocks	drapes	blakes
planes	prunes	pleens
bank	kink	bink
chance	finch	chounce
list	rasp	liced
month	nymph	manth
child	vault	chauld
ships	shacks	shaps

Monosyllabic pseudowords. A set of 24 monosyllabic pseudowords (PSEUDO) was created by scrambling the medial vowels in the high frequency word set. In this way syllabic and phonetic patterns permissible in English phonology were maintained. The frequency of occurrence for these patterns was held constant for the word and pseudoword stimuli. Four adult speakers of English listened to the pseudoword items and judged whether each stimulus could be an acceptable word in English. Two vowel reassignments were made in accord with this feedback, resulting in the pseudoword stimuli listed in Table 2.

Multisyllabic words. The multisyllabic stimuli (MULTI) were three- and four-syllable nouns, all pronounced with stress on the first syllable. Since it is more difficult to control strictly for phonetic parameters in multisyllabic words, the items were selected to represent an array of syllabic and phonetic constructions. For each syllable length an equal number of high frequency and low frequency words was included, again based on word counts from Carroll et al. (1971). The multisyllabic stimuli are listed in Table 3.

Table 3

Multisyllabic Stimuli

High frequency	Low frequency
basketball	badminton
medicine	marmalade
furniture	refugees
neighborhood	saddlebag
vitamins	vinegar
satellite	silicone
television	dormitory
agriculture	anesthetic
helicopter	honeysuckle
supermarket	salamander
military	malnutrition
kindergarten	gladiators

Stimulus preparation. The stimuli were recorded by a phonetically trained male speaker, with each produced as the final word of a meaningful sentence. The sentences were later digitized at 20,000 samples/sec and each stimulus was excised from the sentence, using the Haskins WENDY waveform editing system. The items were arranged into a fixed random sequence for each set of stimuli and were then recorded onto one channel of a magnetic tape with an inter-stimulus-interval (ISI) of 4 secs. At the same time, a series of pulses to be used for timing purposes was recorded on the second channel of the magnetic tape. A pulse was aligned temporally with the onset of each stimulus item.

Control task. A brief 2000 Hz (100 ms) tone was recorded 24 times in two blocks of 12 trials on one channel of an audiotape. The ISI randomly varied with intervals ranging from 2.5 sec to 5 sec. To enable reaction time measures, a pulse was recorded on the second channel to co-occur with each tone.

Apparatus. The stimuli were replayed on a reel-to-reel tape recorder. One channel, containing the stimuli, was output to the subject and to the experimenter via open-air soft-cushion headphones. The other channel, with the pulses, was connected to the onset trigger of a timer. As each word or pseudoword was produced on the tape recorder, the pulse triggered the counter on the timer. The subject would repeat the stimulus, as rapidly as possible, speaking into a pair of microphones centered in front of the subject. One of the microphones contained a voice key, which would terminate the counter. The resulting reaction time, displayed digitally, was written by the experimenter and was output from a printer. Via the second microphone, the subjects' responses were recorded on audiotape. Transcriptions of the responses were also made during the testing session. The response tapes were listened to later in the day in order to corroborate the transcription and to allow any necessary corrections. The same apparatus was used for the control task.

Procedure. Each child was tested individually in a quiet room for three sessions. The first session included the Woodcock reading tasks and the Peabody Picture Vocabulary test. In the second session, occurring at least a week later, the children were given the hearing screening and the monosyllabic word reaction-time task. The third session, occurring approximately another week after the second, included the multisyllabic word RT task, the monosyllabic pseudoword RT task, and the control task. We elected to present the conditions in a single order that we felt would be easy for third graders to follow.

For the speech stimuli tasks, the subjects were asked to say what they heard as quickly as possible. While speed was encouraged, the children were also instructed to say the words distinctly. Prior to the RT tasks, the subjects practiced repeating words said by the experimenter and then practiced repeating preliminary items on the tape.

For the control task, subjects were instructed for the first twelve trials to say the word /cat/, as rapidly as possible, when a tone was heard. For the second block of twelve trials subjects were told to say /banana/ upon hearing a tone.

Results and Discussion

The responses were analyzed in terms of accuracy (number correct) and speed (reaction time).

Accuracy scores. The responses were scored for phonetic accuracy. Each item was scored as correct or incorrect. If a subject stuttered or stammered during a response, this was not counted as an error. Any other misproduction, changing the phonetic description of the item, was noted as an incorrect response. The results are presented in Figure 1. Since the order of presentation of conditions was not counterbalanced, comparisons between performance of reading groups will be made within each set.

On the monosyllabic words, which we had characterized as the least difficult set, the reading groups performed comparably, $F(1,28)=.79$, $p=.38$. More errors occurred on the low frequency words, $F(1,28)=39.79$, $p<.0001$, but this was true for both reading groups, as can be seen in the lack of a frequency x group interaction, $F(1,28)=.61$, $p=.44$. However, with the more demanding conditions, the poor readers produced significantly more errors. On

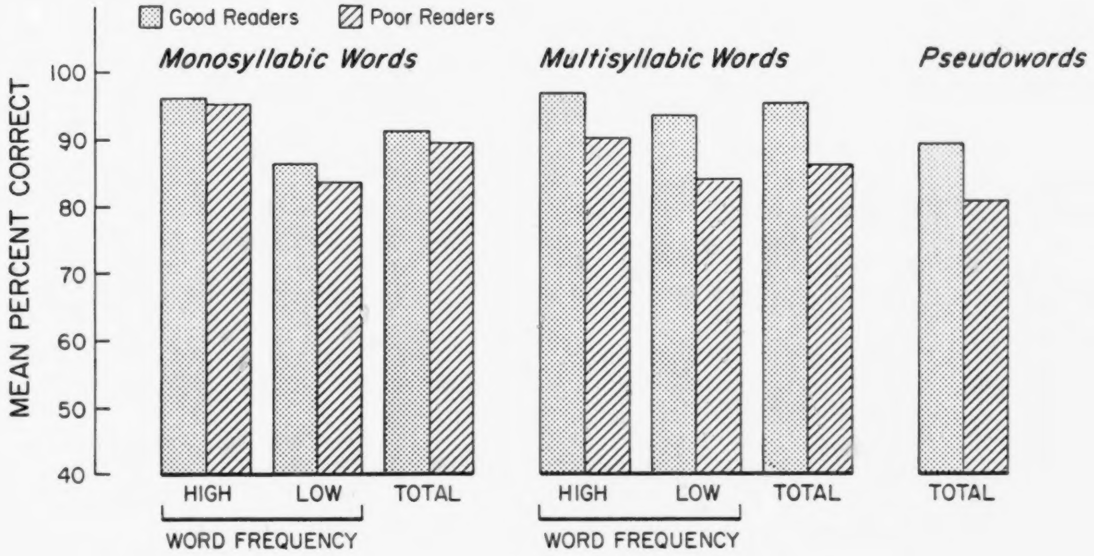


Figure 1. Accuracy performance of good and poor readers, plotted in mean percent correct.

the *multisyllabic* stimuli, group differences were obtained on the entire set, *F*(1,28)=8, *p*=.009, and on both the high frequency, *F*(1,28)=5.49, *p*=.03, and low frequency, *F*(1,28)=6.45, *p*=.02, stimuli. Once again there was an overall effect of word frequency, *F*(1,28)=7.78, *p*=.01, but this did not differ for good and poor readers, *F*(1,28)=.66, *p*=.42. An additional analysis was performed on the MULTI data, examining the effect of the length of the stimuli on the error rate. Both good and poor readers tended to produce more errors on the longer, four-syllable items, though this pattern was not significant, *F*(1,28)=3.43, *p*=.08. While longer utterances may be more difficult to process, the particular phonetic sequence required appears to be a more salient factor. For example, in the four syllable stimuli no errors were obtained on the item /salamander/ while many children mispronounced the cluster in /agriculture/.

Since word frequency effects were obtained on both the MONO and MULTI conditions, one might predict an even higher error rate on the pseudowords, given that subjects obviously have no prior familiarity with these utterances. For poor readers this looks to be the case: they produced the most errors on the pseudoword stimuli. Good readers, on the other hand, had fewer errors on average on the pseudoword stimuli than on the low frequency monosyllabic real words. The good readers appear to have benefited from the previous trials, getting more experienced with the task and perhaps getting more finely tuned to the phonetic requirements of the task (e.g., adjusting to the particular dialect of the speaker).[2] Thus the difference in performance between reading groups widened in the PSEUDO condition, again yielding significant results, *F*(1,28)=9.98, *p*=.004.

In sum, for accuracy measurements a noteworthy difference in performance was observed for the two reading groups. As we had predicted, the poor readers made significantly more errors on the multisyllabic and pseudoword conditions. The comparable effects of word frequency for both reading groups suggests that the perceptual problems of poor readers do not stem from possible differences in word knowledge. Our next step was to check whether IQ level might have been the underlying basis for these reading group results. Although the groups did not significantly differ on PPVT-R IQ scores, Crowder (1984) has pointed out that this may not adequately control for IQ factors. He argues that the size of the obtained group difference in IQ is not relevant in light of regression artifacts that may exist. To address this concern, one can test whether reading group differences in IQ might be responsible for the obtained results by recombining the subjects into high and low IQ groups. When this was done (high IQ: $\bar{x}=113.9$; low IQ: $\bar{x}=98.5$), the conditions that had revealed significant reading group effects were reanalyzed and no significant IQ group differences were evident (MULTI: $F(1,28)=.91$, $p=.35$; PSEUDO: $F(1,28)=.21$, $p=.65$). These results support the conclusion that the findings of speech perception differences for the good and poor readers arise from factors related to reading ability per se.

Analysis of reaction time data. The mean reaction times of correct responses for the three stimuli sets are shown in Table 4. Reaction times are excluded from trials in which the response was incorrect and/or the subject's reaction time was not within the limits of 200-2000 ms.

Table 4

Mean Reaction Time (ms) for Correct Trials

	Monosyllabic Words			Multisyllabic Words			Pseudowords
	High Frequency	Low Frequency	Total	High Frequency	Low Frequency	Total	
Good	847.6	876.7	861.2	824.8	875.7	852.4	732.6
Poor	818.3	857.4	838.8	760.7	788.5	772.1	686.7

As described earlier, the conditions were presented in a single order (1. MONO; 2. MULTI; 3. PSEUDO), which we thought would be easy for third graders to perform. A general observation can be made that RT values get faster for successive blocks of trials (as is typical for adults), overcoming the processing requirements imposed by greater length or reduced word familiarity. However, these effects can be noted within conditions, as will be described below.

The major finding for this scoring procedure is that there are no significant differences in RT between reading groups for any condition: MONO, $F(1,28)=.21$, $p=.65$; MULTI, $F(1,28)=2.80$, $p=.12$; PSEUDO, $F(1,28)=.82$, $p=.37$. Further, although no group differences were significant, we were surprised that the poor readers, rather than being slower than good readers, were on average somewhat faster. This finding will be discussed below in relation to the error data.

With children as subjects, concerns might be raised about the reliability of reaction time data. However, the RT values suggest the subjects were seriously engaged in the task, and the results indicate systematic effects of linguistic parameters. For example, expected word frequency effects (less frequent items taking longer to initiate) were observed on both the monosyllabic lists, $\underline{F}(1,28)=27.63$, $\underline{p}<.0001$, and on the multisyllabic condition, $\underline{F}(1,28)=18.54$, $\underline{p}=.0002$, with no interaction of word frequency with reading groups (MONO: $\underline{F}(1,28)=.59$, $\underline{p}=.45$; MULTI: $\underline{F}(1,28)= 1.59$, $\underline{p}=.22$)

Had RT differences for good and poor readers been evident, we wanted to be able to focus on which aspect of the repetition task might have been responsible: identification of the input or articulation of the response. To do this, we administered the control task in which the speech identification process had been eliminated. Obviously, given the lack of reading group RT differences, the control results did not serve the original purpose. Nonetheless, the results do corroborate the lack of reading group RT differences in the word repetition tasks (monosyllabic control (/cat/): $\underline{F}(1,28)=1.0$, $\underline{p}=.33$; multisyllabic control (/banana/): $\underline{F}(1,28)=2.31$, $\underline{p}=.14$).

In sum, there is no indication that the efficiency of phonetic processing that is represented in reaction time data differs for good and poor readers. We must consider, then, why the reading groups did not differ in reaction time performance, but did contrast on accuracy scores. Traditionally, these two dependent measures of performance have been viewed as alternative ways of studying the same underlying processes (e.g., Eriksen & Eriksen, 1979; Lappin, 1978; Smith & Spoehr, 1974). However, evidence has been reported recently suggesting that speed and accuracy measures do not always reflect the same aspects of information processing (Santee & Egeth, 1982).

In the present study, speed/accuracy tradeoffs appear to be present for both good and poor readers. In Table 5, it can be seen that in some of the conditions significant negative correlations were obtained between RT and the incidence of errors. Our question is whether poor readers' tendency to have faster RTs might be contributing to the observed reading group error differences. For the monosyllabic condition this issue doesn't arise since the good and poor readers were not distinguished by error rate. In the multisyllabic task, the nonsignificant correlations between RT and errors indicate that other factors are the basis of the error performance. In the pseudoword condition, the two dependent measures were correlated, so an analysis of covariance was conducted on the error data using RT as the

Table 5

Correlations for Measures of Reaction Time and Error Rate

	Monosyllabic Words	Multisyllabic Words	Pseudowords
Good Readers	-.20	+.20	-.55*
Poor Readers	-.65*	-.24	-.48*

*$\underline{p}$ < .01

covariate. Significant reading group differences were still evident, $F(1,27)=9.1$, $p=.006$, again suggesting that while error and accuracy scores in part arise from the same processes, other factors are uniquely contributing to the error scores.

To reiterate, the results indicate that poor readers are less accurate in phonetic processing, but are not slower. It appears that it is necessary to have a somewhat demanding task in order to discern reading group differences in phonetic ability. On the more difficult tasks, omega squared was calculated to determine the proportion of variance accounted for by the accuracy of phonetic processing. The results are as follows: MULTI = .14; PSEUDO = .21. These effect sizes indicate that a fair amount of the performance differences between reading groups can be attributed to phonetic processes in perception.

Conclusion

In this study we looked more closely at good and poor readers' performance in speech perception with nondegraded stimuli in an attempt to explore the basis of poor readers' short-term memory deficits. On repetition tasks, RT and accuracy measures were taken for monosyllabic, multisyllabic, and pseudoword stimuli for third-grade good and poor readers. Although there was no indication of reaction time differences for the reading groups, the good readers were significantly more accurate than the poor readers for the more demanding multisyllabic and pseudoword stimuli.

Our framework has been to consider whether differences in phonetic processing efficiency might be central to short-term memory function, which in turn plays a role in spoken and written language comprehension. Assumptions are being made in this approach that have been generally validated in research on cognitive processes. One is the assumption of a limited-capacity working memory system (Baddeley & Hitch, 1974). Second, within that system sub-processes are assumed to become more automatic with experience and to require less resource allocation (LaBerge & Samuels, 1974). Perfetti (1985) has formalized this approach in his "Verbal Efficiency Theory of Reading Ability" and provides a strong case for the role of the efficiency of lower level processes in language processing, and specifically in reading.

Here we are examining one such lower level process, phonetic skills, to attempt to explicate the nature of the linguistic deficits occurring for poor readers on memory tasks. Given the consistent evidence of a relationship between speed of processing and memory span for adults (Baddeley et al., 1975), as well as developmentally (Case et al., 1982; Hulme et al., 1984), it seemed plausible that the perception and memory deficits of poor readers might stem from reduced efficiency of perceptual processes and, consequently, from limited STM resources. Our results were mixed: the quality of responses was significantly less accurate for the more phonetically demanding stimuli, though somewhat surprisingly the poor readers were not found to be slower at initiating a phonetic response. In a subsequent study (Merlo & Brady, in preparation) this pattern has been replicated. Research by others generally conforms to this picture as well. For somewhat demanding speech tasks (speech-in-noise, Brady et al., 1983; multisyllabic words, Snowling, 1981; phonologically difficult phrases, Catts, 1984; tongue twisters, Merlo & Brady, in preparation), poor readers have repeatedly been observed to produce more errors. On the other hand, reaction time measures for tasks entailing creation of a phonetic representation (e.g., object naming, color naming, digit naming, word naming) have generally not revealed reading group differences in RT unless the stimulus involved orthographic information (Katz & Shankweiler, 1986; Perfetti, Finger, & Hogaboam, 1978; Stanovich, 1981). However, there are some indications that differences in naming speed may be present with younger children or more disabled readers (Blachman, 1981; Denckla & Rudel, 1976a, 1976b; Spring & Capps, 1974). In toto, these findings suggest that the important differences in perceptual operations between good and poor readers rest not with the rate of processing, but with the accuracy of formulating phonetic representations.

To summarize, in the present study we have extended previous observations of inferior perception by poor readers with speech-in-noise to perceptual deficits with clearly presented stimuli. These results strengthen the hypothesis that the memory deficits commonly observed in poor readers for linguistic material may derive from the perceptual requirements of the task, that is, from less efficient encoding of the phonetic items.

References

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Eds.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575-589.

Blachman, B. A. (1981). *The relationship of selected language measures and the McCarthy Scales to kindergarten and first-grade achievement*. Unpublished doctoral dissertation, University of Connecticut.

Brady, S. A., Mann, V. A., & Schmidt, R. (1985). Errors in short-term memory for good and poor readers. *Haskins Laboratories Status Report on Speech Research*, *SR-81*, 85-103.

Brady, S. A., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology*, *35*, 345-367.

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. New York: American Heritage Publishing Co.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386-404.

Catts, H. (1984, November). *Speech production/phonological deficits in reading-disordered children*. Paper presented at the American Speech-Language-Hearing Association, San Francisco, CA.

Chi, M. T. H. (1976). Short-term memory limitations in children: Capacity or processing deficits? *Memory & Cognition*, *4*, 559-572.

Crowder, R. G. (1984). Is it just reading? Comments on the papers by Mann, Morrison, and Wolford and Fowler. *Developmental Review*, *4*, 48-61.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, *89*, 63-100.

Denckla, M. B., & Rudel, R. G. (1976a). Naming of object-drawings by dyslexic and other learning disabled children. *Brain and Language*, *3*, 1-15.

Denckla, M. B., & Rudel, R. G. (1976b). Rapid "automatized" naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*, 471-479.

Dunn, L. M. (1981). *Peabody Picture Vocabulary Test--Revised*. Circle Pines, MN: American Guidance Service, Inc.

Eriksen, C. W., & Eriksen, B. A. (1979). Target redundancy in visual search: Do repetitions of the target within the display impair processing? *Perception & Psychophysics*, *26*, 195-205.

Finkinbinder, R. L. (1973). A descriptive study of the Goldman-Fristoe-Woodcock test of auditory discrimination and selected reading variables with primary school children. *The Journal of Special Education*, *7*, 125-131.

Goldman, R., Fristoe, M., & Woodcock, R. W. (1970). *Goldman-Fristoe-Woodcock test of auditory discrimination*. Circle Pines, MN: American Guidance Service.

Holmes, D. R., & McKeever, W. F. (1979). Material specific serial memory deficit in adolescent dyslexics. *Cortex*, *15*, 51-62.

Hoosian, R. (1982). Correlation between pronunciation speed and digit span size. *Perceptual and Motor Skills*, *55*, 1128.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, *38*, 241-253.

Katz, R. B., & Shankweiler, D. (1986). Repetitive naming and the detection of word retrieval deficits in the beginning reader. *Cortex*.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology*, *32*, 474-484.

LaBerge, D., & Samuels, S. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.

Lappin, J. S. (1978). The relativity of choice behavior and the effect of prior knowledge on the speed and accuracy of recognition. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, NJ: Erlbaum.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex*, *18*, 367-375.

Liberman, I. Y., & Shankweiler, D. (1979). Speech, the alphabet and teaching to read. In L. B. Resnik & P. A. Weaver (Eds.), *Theory and practice of early reading* (pp. 109-134). Hillsdale, NJ: LEA.

Mann, V. A. (1984). Reading skill and language skill. *Developmental Review*, *4*, 1-15.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and wordstrings in relation to reading ability. *Memory & Cognition*, *8*, 329-335.

Merlo, M., & Brady, S. (in preparation). Reading ability and short-term memory: The role of phonetic processing.

Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., Finger, E., & Hogaboam, T. (1978). Source of vocalization latency difference between skilled and less skilled young readers. *Journal of Educational Psychology*, *70*, 730-739.

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*, *20*, 241-248.

Santee, J. L., & Egeth, H. E. (1982). Interference in letter identification: A test of feature-specific inhibition. *Perception & Psychophysics*, *31*, 101-116.

Schwartz, A., & Goldman, R. (1974). Variables influencing performance on speech-sound discrimination tests. *Journal of Speech and hearing Research*, *17*, 25-32.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 531-545.

Smith, E. E., & Spoehr, K. T. (1974). The perception of printed English: A theoretical perspective. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in perception and cognition*. Hillsdale, NJ: Erlbaum.

Snowling, M. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research*, *43*, 219-234.

Spring, C., & Capps, C. (1974). Encoding speed, rehearsal and probed recall of dyslexic boys. *Journal of Educational Psychology*, *66*, 780-786.

Stanovich, K. (1981). Relationships between word decoding speed, general name-retrieval ability, and reading progress in first-grade children. *Journal of Educational Psychology*, *73*, 809-815.

Thompson, B. (1963). A longitudinal study of auditory discrimination. *Journal of Educational Research*, *56*, 376-378.

Vellutino, F. R., Pruzek, R., Steger, J., & Meshoulam, U. (1973). Immediate visual recall in poor and normal readers as a function of orthographic-linguistic familiarity. *Cortex*, *8*, 106-118.

Woodcock, R. W. (1973). *Woodcock Reading Mastery Tests*. Circle Pines, MN: American Guidance Services.

Footnotes

[1] However, it may well be the case that low level difficulties creating a phonetic representation in STM may have consequences on higher processes such as comprehension (cf. Mann, 1984; Perfetti, 1985).

[2] This fits unreported observations in previous research we've conducted that both good and poor readers tend to produce errors at the beginning of a demanding speech task, but good readers show more rapid improvement. It would be interesting in future work to evaluate this aspect of phonological competence specifically for good and poor readers.

PHONOLOGICAL AND MORPHOLOGICAL ANALYSIS BY SKILLED READERS OF SERBO-CROATIAN*

Laurie B. Feldman†

Abstract. Two distinctive properties of Serbo-Croatian, the major language of Yugoslavia, have been exploited as tools in the study of reading. First, most literate speakers of Serbo-Croatian are facile in two alphabets, Roman and Cyrillic. The two alphabet sets intersect and words composed exclusively from the subset of characters that occur in both alphabets can be assigned two phonological interpretations--one by treating the characters as Roman graphemes and one by treating the characters as Cyrillic graphemes. By exploiting the availability of two overlapping alphabets, the nature of phonological codes and how they figure in lexical access has been explored. Second, the inflectional and derivational morphology in Serbo-Croatian is complex, and extensive families of morphologically-related words exist. This complex morphology permits one to investigate how morphological structure is appreciated by the proficient language user. In the present report, results of a series of experiments that investigated phonological and morphological analysis in word recognition tasks by adult readers of Serbo-Croatian are summarized and discussed in terms of a characterization of skilled reading in Serbo-Croatian. To anticipate, the skilled reader of Serbo-Croatian appears to appreciate both phonological and morphological components of words.

The Bialphabetic Environment

Serbo-Croatian is written in two different alphabets, Roman and Cyrillic. The two alphabets transcribe one language and their graphemes map simply and directly onto the same set of phonemes. These two sets of graphemes are, with certain exceptions, mutually exclusive. Most of the Roman and Cyrillic letters occur only in their respective alphabets. These are referred to as unique letters. There are, however, a limited number of letters that are shared by the two alphabets. In some cases, the phonemic interpretation of a shared letter is the same whether it is read as Cyrillic or as Roman; these are referred to as common letters. In other cases, a shared letter has two phonemic interpretations, one by the Roman reading and one by the Cyrillic reading; these are referred to as ambiguous letters[1] (see Figure 1). Whatever their category, the individual letters of the two alphabets have phonemic

---

*To appear in A. Allport, D. MacKay, W. Prinz, and E. Scheerer (Eds.) Language Perception and Production London: Academic Press.

†Also University of Delaware

Acknowledgment. This research was supported by funds from the National Academy of Sciences and the Serbian Academy of Science to Laurie B. Feldman; by NICHD Grant HD-01994 to Haskins Laboratories and by NICHD Grant HD-08495 to the University of Belgrade.

[HASKINS LABORATORIES: Status Report on Speech Research SR-85 (1986)]

Table 1

Types of Letter Strings and Their Lexical Status

COMPOSITION OF LETTER STRING	PHONEMIC	INTERPRETATION	MEANING
AMBIGUOUS and COMMON[1]			
BETAP	Roman	/betap/	meaningless
	Cyrillic	/vetar/	wind
POP	Roman	/pop/	priest
	Cyrillic	/ror/	meaningless
POTOP	Roman	/potop/	flood
	Cyrillic	/rotor/	motor
PAJOC	Roman	/pajotc/	meaningless
	Cyrillic	/rajos/	meaningless
COMMON			
MAMA	Roman	/mama/	mother
	Cyrillic	/mama/	mother
TAKA	Roman	/taka/	meaningless
	Cyrillic	/taka/	meaningless

COMPOSITION OF LETTER STRING	PHONEMIC	INTERPRETATION	MEANING
UNIQUE and COMMON[2]			
VETAR	Roman	/vetar/	wind
	Cyrillic	impossible	
ПОП	Roman	impossible	
	Cyrillic	/pop/	priest
ROTOR	Roman	/rotor/	motor
	Cyrillic	impossible	
ПОТОП	Roman	impossible	
	Cyrillic	/potop/	flood
RAJOS	Roman	/rajos/	meaningless
	Cyrillic	impossible	
ПАЈОЦ	Roman	impossible	
	Cyrillic	/pajots/	meaningless

[1] Phonologically bivalent letter strings

[2] Phonologically unequivocal controls

Adapted with permission of the American Psychological Assocation from Feldman and Turvey, 1983 interpretations (classically defined) that are virtually invariant over letter contexts. (This reflects the phonologically shallow nature of the Serbo-Croatian orthography.) Moreover, all the individual letters in a string of letters, be it a word or nonsense, are pronounced--there are no letters made silent by context (see Feldman & Turvey, 1983; Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela, Savić, Gligorević, Ognjenović, & Turvey, 1978).[2]

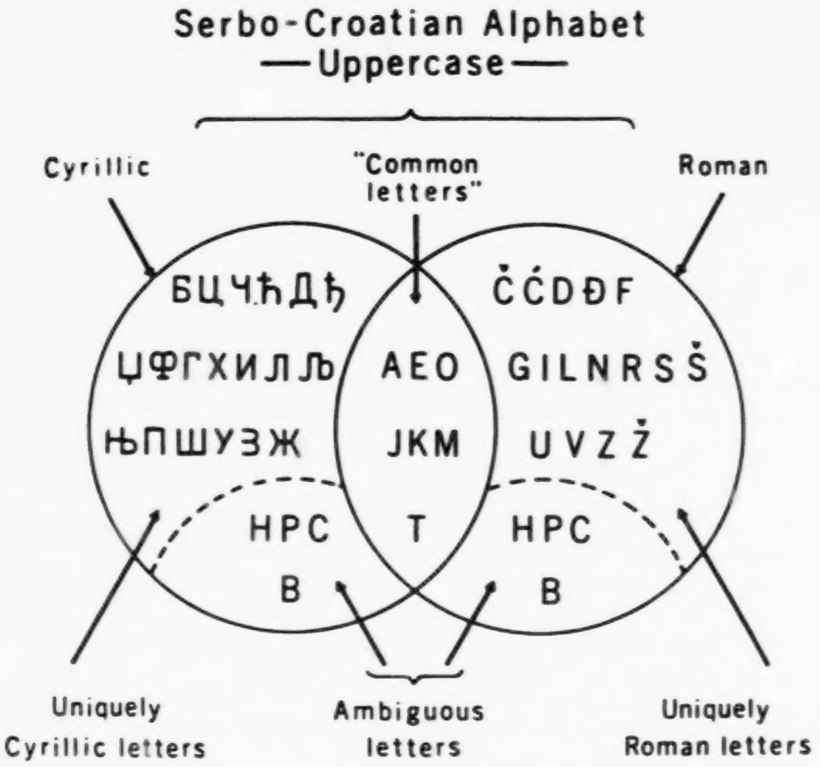

Serbo-Croatian Alphabet
—Uppercase—
Cyrillic
"Common letters"
Roman
БЦЧЋДЂ
ЏФГХИЛЉ
ЊПШУЗЖ
AEO
JKM
ČĆDĐF
GILNRSŠ
UVZŽ
HPC
B
T
HPC
B
Uniquely Cyrillic letters
Ambiguous letters
Uniquely Roman letters

Figure 1. The characters of the Roman and Cyrillic Alphabets (printed from Feldman and Turvey, 1983, with permission from the American Psychological Association).

Given the relation between the two Serbo-Croatian alphabets, it is possible to construct a variety of types of letter strings. A letter string that contains at least one uniquely Roman character in addition to shared characters would be read in only one way and it could be either a word or meaningless. A letter string composed entirely of common and ambiguous letters is bivalent. That is, it could be pronounced in one way if read as Roman and pronounced in a distinctly different way if read as Cyrillic; moreover, it could be a word when read in one alphabet and meaningless when read in the other or it could represent two different words, one in one alphabet and one in the other, or finally it could be meaningless in both alphabets (see Table 1).

Consider the word that means WIND. As with any word in Serbo-Croatian, it can be written in either Roman characters or Cyrillic characters. In its Roman transcription (i.e., VETAR), the word includes unique and common characters and is phonologically unequivocal. By contrast, in its Cyrillic transcription (i.e., BETAP), the word includes only ambiguous and common characters and therefore is phonologically bivalent. By its Cyrillic reading it is a word; by its Roman reading it is meaningless. In the present series of experiments, two forms of the same word are compared where one is phonologically bivalent and the other is phonologically unequivocal. Notice that by comparing two printed forms of the same word, problems of equating familiarity, richness of meaning, length and number of syllables are eliminated.[3] To reiterate, the letter strings exemplified by BETAP and VETAR are the same word and, therefore, identical in all respects but one, namely, the number of phonological interpretations.

Phonological Analysis in Skilled Readers

When bi-alphabetic adult readers of Serbo-Croatian performed a lexical decision task (i.e., Is this letter string a word by either a Roman or a Cyrillic reading?), single letter strings composed of ambiguous and common characters (i.e., those letter strings that could be assigned both a Roman and a Cyrillic alphabet reading) typically incurred longer latencies than the phonologically unequivocal alphabet transcription of the same word. This outcome has been reported both in a mixed alphabet context where the lexical interpretation of a letter string was sometimes in Roman and sometimes in Cyrillic (Feldman & Turvey, 1983; Lukatela et al., 1980) and a pure alphabet context where the lexical interpretation was always in Roman (Feldman, 1983; Lukatela et al., 1978). The effect of phonological ambiguity was significant both for bivalent words and pseudowords, but it was more robust for words. In characterizing the effect of ambiguity in lexical decision, several outcomes prove essential. First, the effect of phonological ambiguity did not vary as a function of word familiarity. For each word, decision latency to its phonologically unequivocal form was used an an index of familiarity and was correlated with the difference in decision latency between the bivalent and unequivocal forms of the word. In lexical decision, that correlation approached zero (Feldman & Turvey, 1983).[4] Second, words composed entirely of common letters (with no ambiguous or unique letters) such as MAMA were accepted as words no more slowly than letter strings that included common and unique letters. Likewise, pseudowords composed entirely of common letters such as TAKA were rejected as words no more slowly than letter strings that included common and unique letters. Note that the distinction between common and ambiguous letters derives from their phonology: each type of letter occurs in both alphabets but only the latter have two phonemic interpretations. The foregoing discrepancy of outcomes suggests that it is phonological bivalence rather than a visually-based alphabetic bivalence that governs the slowing of decision latencies (see Lukatela et al., 1978, 1980, for a complete discussion). Third, lexical decision latencies to letter strings composed entirely of ambiguous and common letters were always slowed whether both alphabet readings yielded a positive response such as "POTOP" (Lukatela et al., 1980) or a negative response such as "PAJOC" (Feldman, 1981; Lukatela et al., 1978, 1980) or the Cyrillic reading and the Roman reading yielded opposite responses such as "BETAP" or "POP" (Feldman & Turvey, 1983; Lukatela et al., 1978, 1980). This outcome invalidates a decision stage account of the detriment due to bivalence that posits some type of post-lexical interference between conflicting lexical judgments. Moreover, insofar as lexical decision is alleged to be susceptible to decision-stage influences in a way that naming is not (Balota & Chumbley, 1984; Seidenberg, Waters, Sanders & Langer, 1984), it is noteworthy that the detriment due to bivalence is generally enhanced in naming relative to lexical decision. Finally, the difference in decision latency between the bivalent and unequivocal forms of a word increased as the number of ambiguous (but not common) characters increased (Feldman & Turvey, 1983). It was eliminated, however, by the presence of a single unique letter (Feldman, Kostić, Lukatela, & Turvey, 1983). These findings imply that a segmental phonology is assembled from an analysis of a letter string's component orthographic structure and that sometimes (multiple) phonological interpretations are generated. The foregoing results of lexical decision experiments with phonologically bivalent letter strings provide evidence that access to the lexicon in Serbo-Croatian necessarily involves an analysis that 1) is sensitive to phonology and component orthographic structure, 2) is not sensitive to the lexical status of the various alphabetic readings. These results have been interpreted as evidence for an assembled segmental phonology in Serbo-Croatian.

In an attempt to understand conditions under which phonological codes and lexical knowledge do interact in Serbo-Croatian, we have begun to explore associative priming of phonologically bivalent words (Feldman, Lukatela, Katz, & Turvey, forthcoming). In this procedure, target words are sometimes presented in the context of another word that is associated with it and decision latencies to the target with and without its associate are compared. Phonologically bivalent words and the unequivocal alphabet transcription of those same words were presented as targets in a lexical decision task. Half of the bivalent targets were words by the Cyrillic reading and half were words by their Roman reading. On some proportion of trials, target words were presented in the context of another word that was associatively related to it and preceded it by 700 ms. Sometimes, the alphabet of the associate was congruent with the alphabet in which the target reading was a word. Sometimes the associate and the target reading were alphabetically incongruent. Results showed significant facilitation in the context of associates, evidence of lexical mediation. More interestingly, decision latencies for bivalent letter strings that are words by one of their alphabet readings were reduced less when those words are preceded by an associate printed in the other, incongruent alphabet than when the associate was printed in the same alphabet as the word reading of the target. This outcome suggests alphabetic congruency as a second source of facilitation. For example, bivalent BETAP, which means WIND when read as Cyrillic, was preceded by the word for STORM. Inspection of word means in Table 2 reveals that target decision latencies for BETAP type words were 64 ms faster when preceded by the Cyrillic form of the word for STORM than by the Roman form of the same word. By contrast, target decision latencies for the same words written in their phonologically unequivocal form were facilitated equally by the prior presentation of an associated word printed in either the congruent or incongruent alphabet. For example, WIND written in Roman, namely VETAR, is phonologically unequivocal and decision latencies were not significantly different when the word for STORM appeared in its Cyrillic or Roman form. Likewise for pseudowords, alphabet congruency had no effect (see Table 2).

Table 2

Lexical Decision (ms) to Bivalent Words and their Unequivocal Controls in the Context of Alphabetically Congruent and Alphabetically Incongruent Associates

ALPHABET OF ASSOCIATE:	BIVALENT (BETAP)	UNEQUIVOCAL (VETAR)
CONGRUENT	709	692
INCONGRUENT	775	685
(NO ASSOCIATE)	845	765

From Feldman, Lukatela, Katz, and Turvey (in preparation)

In summary, lexical decision latencies for phonologically bivalent letter strings are reduced significantly more when preceded by associates that are alphabetically congruent with the word reading of the letter string, than by associates that are not congruent. By contrast, decision latencies for phonologically unequivocal letter strings are not influenced by the alphabet of the associate. Associative and alphabetic sources of facilitation can be identified. Whereas facilitation by association occurs for all the words and is assumed to be lexical in origin, facilitation by alphabet congruency of associate and target was important only for bivalent letter strings. The special dependency of alphabetic congruency on ambiguity suggests that alphabetic priming and phonological ambiguity have a common origin.

In summary, studies of phonological ambiguity indicate that skilled readers of Serbo-Croatian analyze words phonologically. In judging letter strings composed exclusively of ambiguous and common letters for a lexical decision, adult readers appear to assign a phonological interpretation (or several) to each character (Feldman & Turvey, 1983). At the same time, the alphabet in which a prior occurring associate is printed appears to bias the generation or the evaluation of various phonological interpretations of a bivalent letter string. An analogous effect is absent in phonologically unequivocal words and in all pseudowords.

Morphological Analysis in Skilled Readers

The effect of phonological ambiguity has provided a means to evaluate the analytic skills of readers with respect to morphological components. As noted above, the Serbo-Croatian language, in a manner that is characteristic of Slavic languages generally, makes extensive use of inflectional and derivational morphology. A noun can appear in any of seven cases in the singular and in the plural where the inflectional affix varies according to its gender, number, and case. For example, the words STAN and KORA, which mean "apartment" and "crust," respectively, in nominative case can be inflected into six other cases in the singular and in the plural, and different inflectional affixes mark each case (with some redundancy of affixes). Similarly, derived forms for "little apartment" or "thin crust" can be generated by adding one of the diminutive affixes (viz., CIĆ ICA, ENCE, AK) to the base word to produce STANCIC and KORICA, respectively. The prevalence of inflectional and derivational formations in Serbo-Croatian is evidence of its productiveness (see Table 3).

Table 3

Examples of Morphologically-related Words Formed with the Base Morpheme "PIS" Meaning "write"

EXAMPLE	DERIVATIONAL PREFIX	BASE MORPHEME	DERIVATIONAL SUFFIX	INFLECTIONAL SUFFIX	MEANING*
OPIS	O	PIS			description
OPISI	O	PIS		I	descriptions (nom. plural)
PIŠEM		PIŠ		EM	I write (1p. sing)
PIŠETE		PIŠ		ETE	you write (2p. plural)
PISAC		PIS	AC		writer
PISCIMA		PIS	C	IMA	writers (dat. plural)
PISMO		PIS	MO		letter
POPIS	PO	PIS			inventory
POTPIS	POT	PIS			signature
SPISAK	S	PIS	AK		list

*all words are in nominative singular unless otherwise noted

One way in which sensitivity to morphological constituents is construed is in terms of a morphological parser that operates prior to lexical access such that affixes are stripped from a multimorphemic word and the base morpheme serves as the primary unit for lexical search (see Caramazza, Miceli, Silveri, & Laudanna, 1985). Frequency of the base unit and the whole word as well as the difficulty of segmenting the appropriate base unit figure significantly in decision latency (Taft, 1979; Taft & Forster, 1975). In one experiment (Feldman et al., 1983) the effect of phonological ambiguity was exploited to assess whether the base morpheme or the whole word serves as the unit for lexical access of inflected words in Serbo-Croatian. Words were presented in nominative and dative case for a lexical decision. Words were selected so that the nominative case and the base morpheme (i.e., nominative minus inflectional affix for most singular nouns) were phonologically bivalent in the Cyrillic alphabet and phonological unequivocal in Roman. For example, the nominative case of the word meaning VEIN is composed entirely of ambiguous and common letters when printed in Cyrillic (i.e., BEHA) and is therefore phonologically bivalent. In Roman, by contrast, it comprises unique and common letters (i.e., VENA) and is, therefore, phonologically unequivocal. Importantly, in the dative case, neither alphabet rendition is bivalent because the inflectional affixes for words of its class are the phonemes /u/ and /i/, both of which are represented by a unique letter in each alphabet, although the base morpheme of the Cyrillic form (i.e., BEH) is still bivalent.

The major outcome of that experiment was a significant interaction of alphabet and case. The difference in latency between dative nouns presented in Cyrillic and Roman was -28 ms which was not significant, whereas the difference between nominative nouns was 304 ms, which was significant. In that dative nouns always included a unique letter, it appears that the effects of phonological bivalence do not occur if letter strings composed of ambiguous and common characters contain even one unique character. Importantly, in that experiment, the unique character always constituted an inflectional morpheme. Stated in terms of morphological units, the outcome of that experiment was that an inflectional affix composed of a unique character and appended to a bivalent base morpheme canceled the detriment due to ambiguity. Evidently, the reader could use the alphabet designation of the inflectional affix to assign a reading to the base morpheme. In conclusion, bivalence defined on the word but not on the base morpheme alone slowed performance on a lexical decision task. This outcome indicates that lexical access of inflected nouns is not restricted to information in the base morpheme unit. Rather, it encompasses the entire word.

An alternative perspective on a reader's appreciation of morphology assumes that lexical entries are accessed from whole word units and that the principle of organization among lexical entries or the lexical representations themselves capture morphological structure. The final experiment (Feldman & Moskovljević, in press) exploits the complex derivational morphology of Serbo-Croatian to provide further evidence that whereas the morphological structure of words is accessible to the skilled reader, lexical entries are not accessed from base morphemes. The experiment incorporated a comparison of three types of nouns all in nominative case: (1) base forms (e.g., STAN, KORA); (2) the diminutive form of those same nouns, which as described above, is formed (productively) by adding one of the suffixes ČIĆ, ICA, ENCE, AK to the base form (e.g., STANČIĆ, KORICA), where choice of suffix is constrained by gender of the noun, and (3) an unrelated monomorphemic word whose construction inappropriately suggests that it contains the same base form and a diminutive affix (e.g., STANICA, KORAK). The latter are referred to as pseudodiminutive nouns. The examples mean "station" and "step," respectively.

The experimental design was a variation on the primed lexical decision task borrowed from Stanners and his colleagues (Stanners, Neiser, Hernon, & Hall, 1979) and known as repetition priming. In the present adaptation of the task, base forms appeared as target words preceded 7 to 13 items earlier in the list by a prime that was either the identical word again in its base form, its diminutive or a pseudodiminutive form. Decision latency to the target as a function of which type of prime preceded it was examined. In addition, decision latencies to the first presentation of the word in its base, diminutive, and pseudodiminutive forms were compared. Results are summarized in Table 4.

Table 4

Lexical Decision (ms) to Target Words Preceded by Identity, Dimunitive, or Pseudodiminutive Primes

PRIME		TARGET		TYPE OF PRIME
STAN	610	STAN	563	IDENTITY
STANČIĆ	754	STAN	585	DIMINUTIVE
STANICA	718	STAN	609	PSEUDODIMINUTIVE

From Feldman and Moskovljević (in press)

Decision latencies on primes were fastest for base forms, followed by pseudodiminutives and lastly, diminutives. The pattern corresponded with that predicted by frequency and provided no evidence that monomorphemic pseudodiminutive forms were slowed by an inappropriate parsing of morphemic structure. In addition, latencies for base and diminutive forms correlated significantly and neither correlated with pseudodimimutive forms. An examination of target latencies provided further evidence that pseudodiminutive words are not associated with an inappropriate base morpheme (and affix), whereas true morphological relationships are appreciated. Decision latencies to target words that were preceded by pseudodiminutive words were as slow as target words presented for the first time. In contrast, both base word and diminutive primes significantly reduced target decision latencies. In summary, results in the repetition priming variation of lexical decision showed significant facilitation for morphological relatives and no facilitation for unrelated pseudodiminutive words. In light of the claim that semantic relatedness of prime to target does not facilitate target decision latencies at lags as long as those introduced in the present task (Dannebring & Briand, 1982; Henderson, Wallis, & Knight, 1984), the foregoing results are interpreted as morphological in nature. In conclusion, the present experiment showed that the skilled reader of Serbo-Croatian is sensitive to morphological structure as evidenced by the results in repetition priming, but offered no evidence that morphological analysis entails decomposition to a base morpheme prior to lexical access.

In summary, an examination of results from lexical decision and naming tasks that take advantage of the bi-alphabetic condition in Serbo-Croatian provides evidence that skilled reading in Serbo-Croatian proceeds with reference to phonology. Specifically: 1) Skilled readers are slowed when a letter string is phonologically bivalent relative to when it is phonologically unequivocal. 2) The alphabet congruency of a prior-occurring associate can speed decision latencies for phonologically bivalent (but not unequivocal) words. Moreover, it appears that phonological bivalence is defined on the entire word, not in the base morpheme alone, which suggests that 3) Skilled readers do not attempt lexical access from an isolated base morpheme. Concurrently, they consider its affix. Failure to find evidence that base morphemes are the units for lexical access should not be construed as a claim against morphological analysis by the reader, however. The results from repetition priming indicate that prior presentation of a morphological relative but not of a visually similar word facilitates decision latency to a target. The foregoing results support the claim that the skilled reader of Serbo-Croatian analyzes words both phonologically and morphologically.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340-357.

Caramazza, A., Miceli, G., Silveri, C., & Laudanna, A. (1985). Reading mechanisms and the organisation of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology*, *2*, 81-114.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology*, *36*, 435-444.

Diringer, D. (1948). *The alphabet: A key to the history of mankind*. London: The Fleet Street Press.

Feldman, L. B. (1981). Visual word recognition in Serbo-Croatian is necessarily phonological. *Haskins Laboratories Status Report on Speech Research*, *SR-66*, 167-202.

Feldman, L. B. (1983). Bi-alphabetism and word recognition. In D. Rogers & J. A. Sloboda (Eds.), *The acquisition of symbolic skills*. New York: Plenum.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. *Psychological Research*, *45*, 55-72.

Feldman, L. B., Lukatela, G., Katz, L., & Turvey, M. T. (in preparation). Alphabetic and associative priming of phonologically ambiguous words.

Feldman, L. B., & Moskovljević, J. (in press). Repetition priming is not purely episodic in origin. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Feldman, L. B., & Turvey, M. T. (1983). Visual word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 288-298.

Henderson, L., Wallis, J. & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouhuis (Eds.), *Attention and performance X*. London: Erlbaum.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, *8*, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetical lexical decision. *Language and Speech*, *21*, 142-165.

Magner, T. F., & Matejka, L. (1971. *Word accent in Serbo-Croatian*. University Park, PA: Pennsylvania State University Press.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1-30.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition*, *12*, 315-328.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, *18*, 399-412.

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effects in word recognition. *Memory & Cognition*, *6*, 410-415.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *9*, 263-272.

Taft, M., & Forster, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*, 638-647.

Wellisch, H. H. (1978). *The conversion of scripts--its nature, history, and utilization*. New York: John Wiley & Sons.

Footnotes

[1] The introduction of two alphabets into Yugoslavia reflects the influence of the Orthodox Church in the Eastern regions and the Catholic Church in the Western regions. The Cyrillic script is probably an adaptation of the Greek uncial alphabet of the 9th century A.D. and the Roman script is a variation of the Latin alphabet, which was also derived from the Greek, probably via Etruscan (Diringer, 1948). In both cases, the scripts had to be adjusted to represent sounds not present in the Greek language and several mechanisms have been identified: 1) Combining two or more characters to represent a single phoneme such as DZ and, arguably, LJ and NJ; 2) Adding a diacritical mark to an existing letter to form a new letter such as Č, Ć, Š. The creation of new letters by inclusion of a diacritic is particularly prevalent in the adaptation of Roman script to languages whose repertoire of phonemes differs greatly from the Latin. Palatal-alveolar fricatives and affricates are represented in this fashion in many Slavic languages, including Serbo-Croatian (Wellisch, 1978); 3) Taking an existing symbol that was not used in the new language to represent a phoneme not present (or represented by multiple symbols) in the old language. For example, Roman C became /ts/ and Roman K remained /k/; 4) Borrowing characters from other scripts. Insofar as particular adaptations were made independently in each alphabet and the shape of some letters (e.g., D,S,R) were modified slightly in the transition to Latin (Diringer, 1948), the intersection of the two alphabet sets represents a complex of factors.

[2] One consequence of the consistent mapping of grapheme to phoneme is that many dialectal variations are represented in writing such that spelling as well as pronunciation can vary fron region to region. For example, the word that means MILK is MLEJKO in the dialects near Belgrade and is MLIKO in dialects along the Dalmatian Coast. It is important to note that the orthography fully specifies segmental phonolgy but that accent (rising/falling; long/short) is not represented. While vowel accent may differentiate between two semantic interpretations of a written letter string, this distinction is often ignored, however, especially in the dialects of the larger cities (Magner & Matejka, 1971).

[3] By law, all elementary school students must demonstrate competence to read and write in both alphabets. With the exception of liturgical text, which is relatively uncommon, the choice of alphabet is not systematic across genres of printed material. Therefore, it can be assumed that the Roman and Cyrillic forms of a word are equally familiar to the skilled reader.

[4] In naming, however, more familiar words showed smaller effects of phonological ambiguity (Feldman, 1981). Analogous to claims made from studies with English materials (Seidenberg, 1985; Stanovich & Bauer, 1978), those words that are recognized more slowly and are presumably less familiar are more susceptible to phonological effects in a naming task than are less-familiar words.

VISUAL AND PRODUCTION SIMILARITY OF THE HANDSHAPES OF THE AMERICAN MANUAL ALPHABET*

John T. Richards,† and Vicki L. Hanson

Abstract. Two experiments were performed to examine the nature of handshape similarity for the 26 elements of the American manual alphabet. Forty deaf college students, half native (first language) signers of American Sign Language and half nonnative signers, participated in the study. In Experiment 1, subjects were asked to base their judgments on visual characteristics of the shapes. In Experiment 2, they were asked to base their judgments on aspects of manual shape production. Hierarchical clustering and multidimensional scaling analyses showed the two sets of judgments to be quite similar. No clear differences were found between native and nonnative signers in either experiment. These data provide a basis for the future manipulation and detection of manual coding in the processing of verbal stimuli.

In recent years, there has been considerable interest in the cognitive processes of deaf persons (see for example, Conrad, 1979; Furth, 1973; Neville, Kutas, & Schmidt, 1982), frequently focusing on the use of speech-based and manual codes in the processing of verbal materials (Bellugi, Klima, & Siple, 1975; Dodd & Hermelin, 1977; Hanson, 1982a; Quinn, 1981; Treiman & Hirsh-Pasek, 1983). Experimentation in this area often requires an understanding of stimulus similarity so that confusability and selective interference can be systematically varied (see, e.g., Hanson, Liberman, & Shankweiler, 1984; Locke & Locke, 1971). Although several studies have characterized the phonetic similarity of common stimuli (e.g., Miller & Nicely, 1955, for English consonants; Conrad, 1964, for letter names), an adequate characterization of comparable manual stimuli has not been done.

Two different forms of manual language are used by deaf individuals in the course of conversation: Fingerspelling and sign. Fingerspelling, like spoken languages, uses temporal sequencing of constituent elements to convey morphemes. The handshapes shown in Figure 1 constitute these elements. Each is a one-handed representation of a letter in the American Manual alphabet.

---

*Perception & Psychophysics, 1985, 38, 311-319.

†IBM Thomas J. Watson Research Center

Acknowledgment. This work was supported, in part, by grant NS-18010 from the National Institute of Neurological and Communicative Disorders and Stroke. We thank John Conti for his drawings of handshapes used in the first experiment and Nancy Fishbein for testing the subjects.

Words are spelled out by producing these handshapes sequentially in the space to the side of the signer's face.[1] Although many of the shapes are similar to those used in sign, fingerspelling does not use the other parameters essential to sign language in making linguistic distinctions (Klima & Bellugi, 1979; Stokoe, Casterline, & Croneberg, 1965). Fingerspelling is used to convey specific names or words for which no sign equivalent exists and can be used to convey entire conversations (the Rochester method).

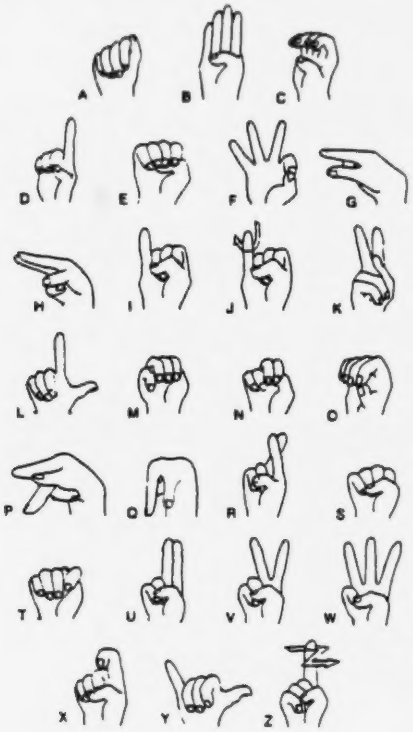


Figure 1. Drawings used as stimuli in Experiment 1.

The present paper focuses on the visual and production similarity of the 26 elements of the American manual alphabet. Deaf college students, both native (first language) and nonnative signers of American Sign Language (ASL), served as informants. Previous studies had examined only subsets of these handshapes (Lane, Boyes-Braem, & Bellugi, 1976; Locke, 1970; Stungis, 1981), or had used hearing subjects with limited prior fingerspelling experience (Weyer, 1973).[2] Experiment 1 examined the similarity of handshapes as visual objects. Experiment 2 examined production similarity.

Experiment 1

Method

Stimuli. Simple line drawings of the 26 handshapes of the American manual alphabet were the stimuli in this experiment. These handshapes and the letters they represent are shown in Figure 1 (note that the letters did not appear with the experimental stimuli). Each handshape was individually rendered on a card measuring approximately 2 1/2 by 3 1/4 in.

Procedure. Subjects were tested individually. At the beginning of an experimental session, the 26 cards were laid out in front of the subject. The arrangement was random, with the constraint that each handshape appear in the proper orientation (i.e., the top of each handshape was always to be on the top).

The subjects were instructed to sort the handshapes into piles on the basis of visual similarity. The following written instructions were presented to the subjects: "The 26 letters of the manual alphabet are laid out in front of you. Begin by looking at each handshape and paying attention to how it looks. Then put the handshapes into piles, so that handshapes that look similar are in the same pile. You can have as many piles as you wish and you can have any number of handshapes in each pile. You can change your mind as often as you like until your arrangement seems best." The experimenter, a deaf native signer of ASL, discussed the instructions with each subject in sign to make sure that the task was clearly understood.

Subjects. The subjects for the experiment were 20 prelingually deaf students from Gallaudet College. Half were native signers of ASL (having learned ASL as a first language from their deaf parents) and half were not. The nonnative signers reported a minimum of 13 years' signing experience. On the average, they had learned to sign at the age of 6.2 years; the mean length of signing experience for these subjects was 18.7 years. All subjects were paid for their participation in this 15-min experiment.

Results and Discussion. Table 1 summarizes the number of subjects who sorted a given handshape pair into the same pile. A score of 20 thus represents the maximum possible interitem similarity. To discover any structure inherent in this matrix (to discover, that is, how the handshapes might be naturally grouped), we applied Johnson's (1967) hierarchical clustering procedure (after first converting the raw frequency counts to a dissimilarity matrix). Separate analyses using the maximum and minimum methods for determining intercluster distance were conducted. Johnson observed that the two methods yield very similar results (at least with the sort of data considered in his report). This was true of the present experiment, in which the maximum and minimum results shared 9 of the 10 letter-pair clusters. Johnson also noted that when the results of the maximum and minimum methods diverge, those obtained with the maximum method appear to be more interpretable. In Figure 2, we show the clusterings produced by the Maximum method. In this figure, similarity decreases as one goes from the top to the bottom and clusters are indicated by adjacent x's. Thus M and N can be seen to be more strongly clustered than A and T, which are more strongly clustered than C and O, and so forth.

It can be seen that only one cluster combines an appreciable number of handshapes; A, T, M, N, E, and S form a group characterized by compactness. Most of the remaining clusters--handshape pairs--appear to be grouped on the basis of a single essential similarity: For the pairs K-P, G-Q, I-J, and D-Z, the letters are formed by identical shapes, which differ only in orientation (K-P, G-Q) or movement (I-J, D-Z); the pair C-O represents two degrees of what might be called hand closure, other aspects (orientation, finger configuration) being the same; and the pairs V-W and B-U differ only in the number of fingers extending upward.

Table 1

Number of Subjects Sorting Handshape Pairs Into Same Pile on the Basis of Visual Similarity in Experiment 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
B	2																								
C	5	0																							
D	0	5	0																						
E	15	1	7	0																					
F	0	6	0	4	0																				
G	0	0	0	0	0	0																			
H	0	4	0	1	0	1	8																		
I	0	3	0	8	0	4	0	0																	
J	0	2	0	6	0	2	0	0	18																
K	0	3	0	4	0	7	0	0	4	2															
L	0	4	0	14	0	4	0	1	8	6	4														
M	15	1	4	0	16	0	0	0	0	0	0	0													
N	15	1	4	0	16	0	0	0	0	0	0	0	20												
O	9	0	16	0	11	0	0	0	0	0	0	0	8	8											
P	0	0	0	0	0	1	5	5	0	1	10	0	0	0	0										
Q	0	0	0	0	0	0	17	5	0	1	0	0	0	0	0	6									
R	0	4	0	6	0	4	1	2	5	3	8	6	0	0	0	4	1								
S	15	1	8	0	17	0	0	0	0	0	0	0	13	13	12	0	0	0							
T	17	1	5	0	14	0	0	0	0	0	0	0	16	16	9	0	0	0	14						
U	0	11	0	5	0	5	0	10	4	2	6	5	0	0	0	0	0	6	0	0					
V	0	3	0	4	0	10	0	1	4	2	14	4	0	0	0	5	0	7	0	0	7				
W	0	6	0	5	0	13	0	2	4	2	10	4	0	0	0	2	0	5	0	0	7	15			
X	2	1	7	4	2	1	0	1	2	1	1	4	2	2	5	0	0	3	3	2	2	1	1		
Y	3	1	2	1	3	3	0	0	3	2	1	2	3	3	2	0	0	1	3	3	1	2	2	3	
Z	0	2	0	15	0	1	0	1	5	6	2	10	0	0	0	1	1	4	0	0	3	2	1	4	1

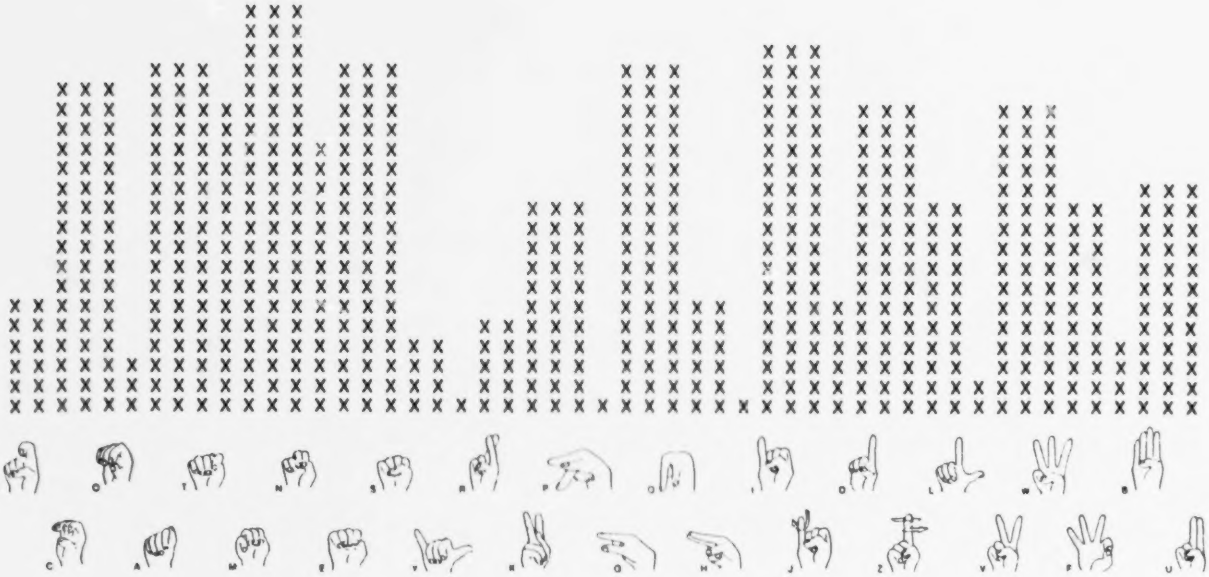


Figure 2. Hierarchical clustering of the handshapes in Experiment 1.

To supplement this cluster-based description, the data were examined for dimensionality and spatially interpretable structure using a nonmetric multidimensional scaling (MDS) procedure (as developed by Kruskal, 1964, and Shepard, 1962). Although a stress analysis suggested no clearly appropriate dimensionality, ease of interpretation leads us to prefer the unrotated two-dimensional solution depicted in Figure 3.

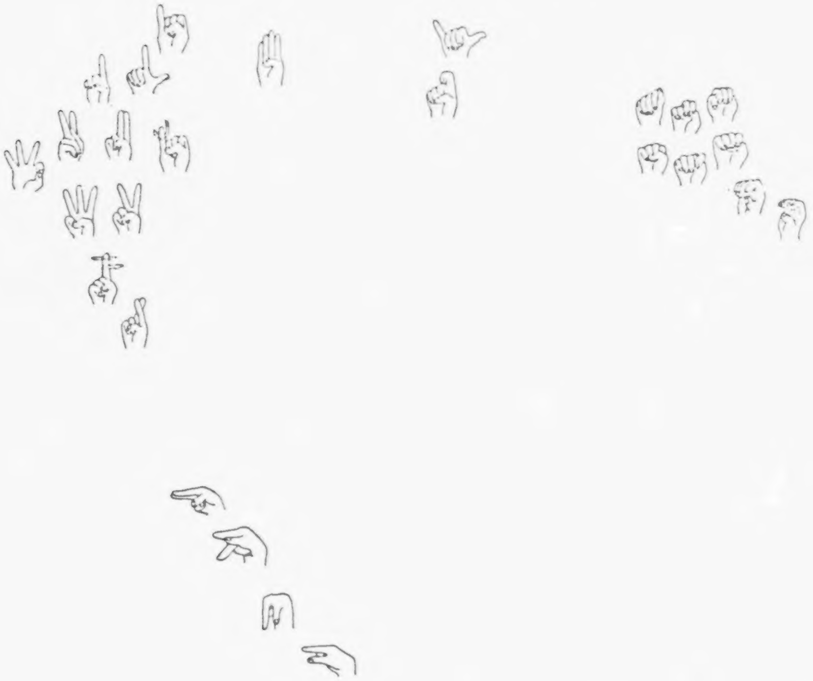


Figure 3. The two-dimensional MDS solution for the handshapes in Experiment 1.

Several aspects of this solution warrant comment. We see the horizontal dimension as representing hand compactness with open or extended handshapes on the left and closed handshapes on the right. The vertical dimension seems best characterized as orientation of the hand's major axis with vertically oriented handshapes near the top and horizontally oriented ones near the bottom. The distribution of the handshapes within this space is somewhat fan like; closed handshapes cluster tightly in the orientation dimension (much more so than can be represented in this figure), whereas open or extended ones are widely dispersed. Another way to say this is that to the extent that closed handshapes have an orientation, it is common to them all.

Finally, an INDSCAL analysis of the individual dissimilarity matrices found no evidence for different organizations as a function of whether ASL was the subject's native language. This may be seen in Figure 4, which plots each subject's weights on the two dimensions of compactness and orientation. There is no evidence that the native signers (filled circles) differ from the nonnative signers (open circles) in their dimensional weightings.

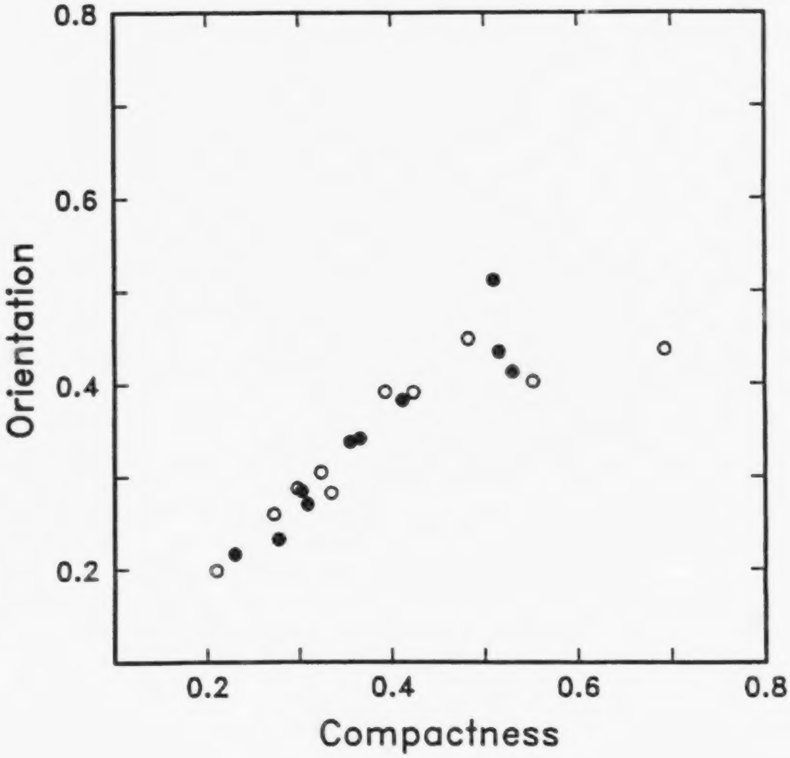


Figure 4. Individual subject's weightings on the two dimensions of orientation and compactness in Experiment 1. Filled circles represent native signers of ASL, open circles represent non-native signers.

Experiment 2

In the second experiment, similarity judgments were based on the essentially kinesthetic aspects of manual handshape production. To help ensure this, uppercase letters were used as stimuli (forcing subjects to generate, either overtly or covertly, the handshapes being compared at any point during the sorting task). Instructions emphasized that production similarity was to be assessed. In other respects, the second experiment was identical to the first.

Method

Stimuli. Stimuli were the uppercase representations of the 26 letters of the alphabet. Each character was printed, in 30-point lettering, on a 3 x 5 in. index card.

Procedure. The procedure was similar to that of Experiment 1. For this experiment, the subjects were instructed to sort the cards into piles on the basis of the similarity of the handshapes. Written instructions were given to the subjects, and a deaf experimenter (the same person as in Experiment 1) reviewed the instructions with subjects to ensure they were understood. The written instructions were as follows: "The 26 letters of the alphabet are laid out in front of you. Begin by thinking about the handshapes for each letter. Then put the letters into piles, so that letters that have handshapes that are similar to produce are in the same pile. You can have as many piles as you wish and you can have any number of letters in each pile. You can change your mind as often as you like until your arrangement seems best. REMEMBER TO THINK ABOUT EACH HANDSHAPE AND GROUP THE LETTERS ACCORDING TO SIMILARITY OF THE HANDSHAPES."

Subjects. Twenty deaf students from Gallaudet College were tested; half were native signers of ASL, the other half were not. The data of one of the nonnative signers were eliminated from analysis due to an apparent failure to follow the instructions (the sorting of this subject was based on the visual similarity of the printed letters rather than on the production similarity of the handshapes, as evidenced by clusters such as W-M, X-K, F-E, and A-H). The remaining nine nonnative signers reported a minimum of 13 years' signing experience. On the average, they had learned to sign at the age of 5.3 years; the mean length of signing experience for these subjects was 17.1 years. All subjects were paid for their participation in this 15-min experiment.

Results and Discussion. Table 2 summarizes the number of subjects who sorted a given handshape pair into the same pile (19 being the maximum possible similarity score). As in the first experiment, these data were subjected to a hierarchical clustering analysis; the result is shown in Figure 5.

In slight contrast to the first experiment, these data appear to possess less global structure. In particular, the compact handshapes (A, T, M, N, E, A, S) exhibit no tendency to cluster as a single group. Rather, two smaller clusters emerge, each being describable in production-relevant terms: The E, A, and S handshapes share the position of the four fingers, differing only in thumb placement relative to the finger group; the M, N, and T handshapes differ only in the number of fingers extended over the thumb, with M having three, N having two, and T having only one. With this exception the remaining clusters--pairs once again--are primarily grouped as before. This similarity between the results of Experiments 1 and 2 is further supported by a moderately large correlation between the matrices shown in Tables 1 and 2 ($r = .66$, $df = 323$, $p < .01$).

Table 2

Number of Subjects Sorting Handshape Pairs Into Same Pile
on the Basis of Production Similarity in Experiment 2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
B	0																								
C	4	0																							
D	1	4	4																						
E	11	2	6	4																					
F	0	7	0	4	3																				
G	0	1	3	0	1	3																			
H	1	2	2	1	3	2	8																		
I	1	1	1	1	1	1	1	1																	
J	0	1	0	0	0	1	1	1	15																
K	0	1	0	0	0	1	0	1	1	0															
L	0	2	0	5	0	2	2	1	8	6	2														
M	3	2	1	1	2	0	1	3	0	0	0	0													
N	2	2	1	1	2	0	1	3	0	0	1	0	18												
O	6	0	12	3	6	0	2	1	1	0	0	0	1	1											
P	0	2	0	2	1	3	1	1	0	0	14	1	0	0	0										
Q	0	0	3	1	0	2	14	5	0	0	0	1	0	0	4	2									
R	0	5	0	3	2	6	3	3	1	1	2	3	0	0	0	3	2								
S	16	0	4	1	10	0	0	1	1	0	0	0	2	2	5	0	0	0							
T	7	0	3	2	5	0	1	2	2	2	0	2	8	8	1	1	0	0	8						
U	0	1	1	0	0	0	1	6	0	0	0	0	2	2	1	0	1	1	0	1					
V	1	1	0	0	0	0	0	2	0	0	0	0	2	1	0	0	0	1	0	1	13				
W	1	2	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	1	11	17			
X	5	0	2	2	3	0	0	1	0	1	0	1	2	1	1	1	0	0	4	4	2	2	2		
Y	2	0	0	1	1	0	0	0	2	3	0	2	2	1	0	1	1	0	1	2	3	4	4	4	
Z	1	1	0	4	1	1	1	2	2	6	1	3	0	1	0	1	0	1	1	3	1	1	1	7	5

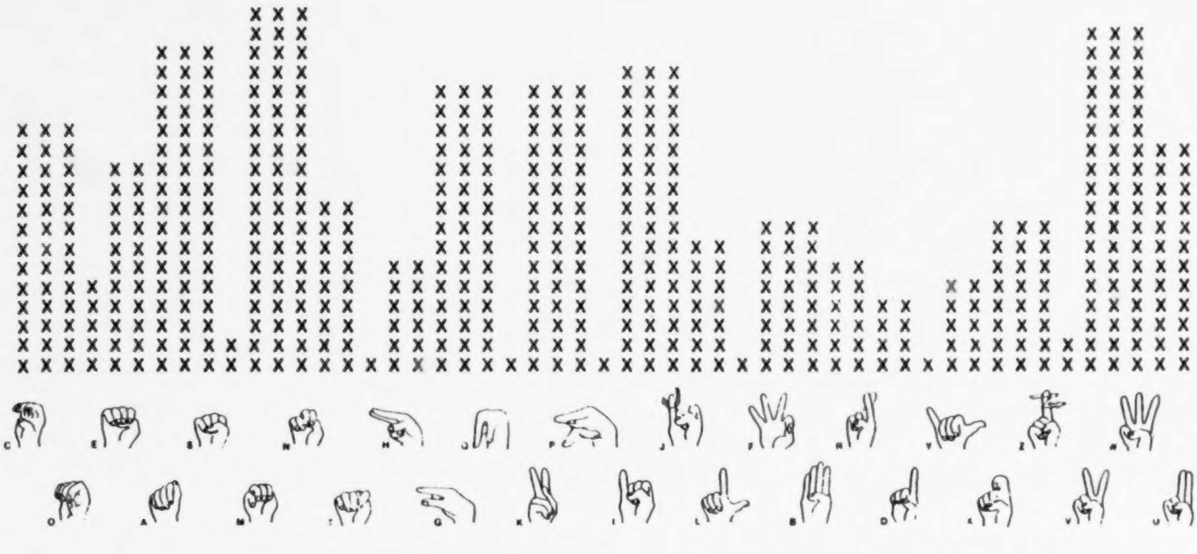


Figure 5. Hierarchical clustering of the handshapes in Experiment 2.

The dimensionality and spatial structure of these data were next analyzed using an MDS procedure. The two-dimensional solution shown in Figure 6 exhibits many of the same characteristics as before. The horizontal dimension represents hand compactness with open or extended handshapes on the left and closed handshapes on the right. The vertical dimension represents orientation of the hand's major axis with vertically oriented handshapes near the top and horizontally oriented ones near the bottom. And although the distribution of the handshapes within this space is somewhat more uniform than in Experiment 1, we view the two solutions as essentially similar.

Further evidence of this similarity derives from a comparison with the solution obtained by Weyer (1973) for visual handshape confusability. Although Weyer did not choose to interpret the two dimensions of his solution, they correspond to the dimensions of compactness and orientation found here. Moreover, the distribution of handshapes within the space is very similar to the distribution shown in Figure 6 (with the only significant exception being a left-right reflection of the compactness dimension). We conclude from this that production similarity and visual similarity are structurally similar.

Finally, we found no evidence for different organizations as a function of whether ASL was the subject's native language. An INDSCAL analysis suggested, as before, that the groups were similarly dispersed within the space of dimensional weights. This is shown in Figure 7.

General Discussion

In the two experiments reported here, visual and production similarity for the handshapes of the American manual alphabet were determined to be essentially the same. For both sets of judgments, the dimensions of hand compactness and orientation were found to describe the data. And for both sets of judgments, similar numbers and arrangements of handshape clusters emerged. We conclude from this that judged handshape similarity is relatively unaffected by the modality to which the judge attends. The present results also suggest that at least within the range of relatively skilled signers, perceived handshape similarity does not vary as a function of degree of experience with fingerspelling; we found no differences between native and nonnative deaf signers.

The present data are quite consistent with earlier studies of perceptual confusability. They are in accord with results reported for the subset of manual alphabet handshapes included in the ASL studies of Lane et al. (1976) and Stungis (1981). In these two studies, the major differentiating feature was whether fingers were extended (open) or not extended (compact).

The present data are also in accord with the results obtained by Weyer (1973) for the entire manual alphabet. Weyer investigated the confusions that emerged during tachistoscopic recognition of computer-generated handshapes. His clustering analysis indicated that the largest cluster was composed of the N, S, T, and A handshapes, with an adjacent cluster composed of the E, M, and O handshapes. These handshapes, characterized by Weyer as involving fists and folding fingers, are the same as those found by our analyses to be "compact." Some of the smaller clusters found by Weyer were also apparent in the visual similarity data of the present Experiment 1, e.g., B-U, V-W, and I-J. Some differences did arise, however, in specific clusterings of handshape pairs.

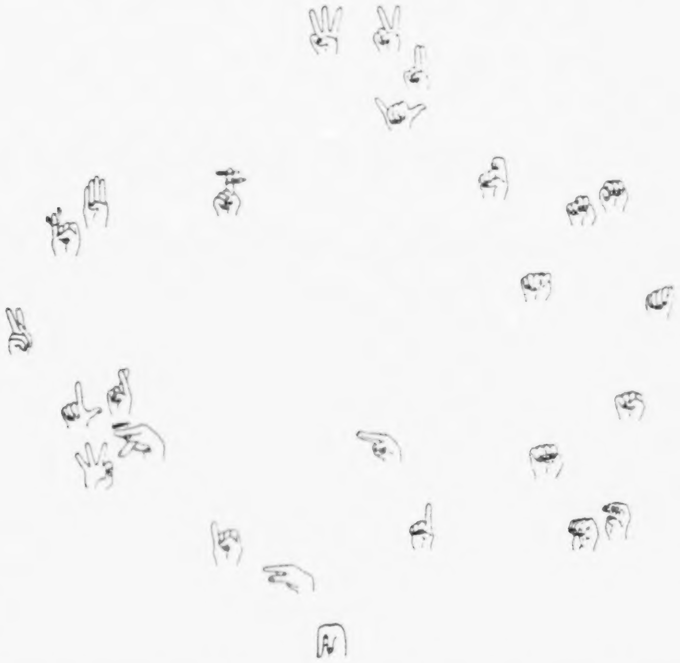


Figure 6. The two-dimensional MDS solution for the handshapes in Experiment 2.

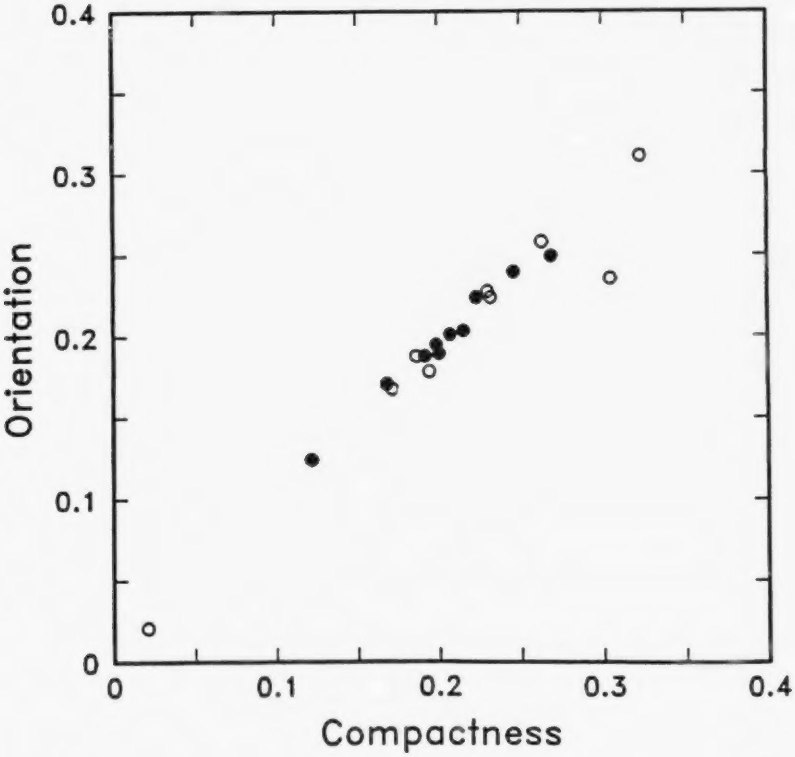


Figure 7. Individual subject's weightings on the two dimensions of orientation and compactness in Experiment 2. Filled circles represent native signers of ASL, open circles represent non-native signers.

We found, for example, that our deaf subjects judged, as visually similar, pairs that had similar shapes but differed in orientation (K-P, G-Q) or movement (D-Z). These groupings were not obtained by Weyer. Such differences might be attributed to procedural variation (Weyer used tachistoscopic recognition; we used a sorting task), or to differences in the handshape stimuli used, or to subject differences. To the extent that the differences are reliable, we suspect that subjects' differing familiarity with the manual alphabet underlies them. Twelve of the 15 subjects in Weyer's study were hearing, and the level of fingerspelling expertise was given for none of the subjects. It is possible that his hearing subjects were totally unfamiliar with fingerspelling prior to the experiment. If so, they would have tended to rely on visual features, whereas our more experienced subjects might have allowed their knowledge of handshape production (e.g., K and P are the same handshape, just oriented differently) to influence their judgments.

The present data are consistent, moreover, with patterns of interletter confusion obtained in tasks requiring the short-term retention of printed letter strings. Two studies, one by Conrad and Rush (1965) and another by Wallace and Corballis (1973), examined short-term retention by deaf subjects with manual language experience (and published the raw confusion matrices needed here). Of these two, only the one by Wallace and Corballis included, in the stimulus set, a high proportion of letters found by our techniques to be manually similar.[3] From this fact alone we might expect that the confusion data of Conrad and Rush would be less influenced by manual similarity than the data of Wallace and Corballis. The correlations summarized in Table 3 are in line with this expectation (note that the results in this table were derived by correlating the interletter confusion matrices, collapsed across conditions within each of the two studies, with the subset of our manual similarity matrices containing the letter subset used in each of the two studies). We find an interpretable pattern of correlations within the conditions of the Wallace and Corballis study as well. In Table 4, separate correlations are shown for stimulus strings of length 4 and 5 for subjects with manual training and for those with oral training. The higher correlations for the longer stimulus strings may well correspond to a greater reliance on language codes in short-term memory. The higher correlations for the manual subject group may well reflect a greater tendency to associate the printed letter strings with the corresponding handshapes (a tendency made all the more likely by their history of instruction in the Rochester method--a technique in which all words are fingerspelled). These two trends are even more apparent in the right half of the table. Here we show the correlations between confusion and similarity matrices from which the letter pair G-Q has been excluded. Since Wallace and Corballis noted that the lowercase forms of their stimulus letters G and Q were highly similar visually (differing only in a right- versus left-hooking descender), and since these two letters are also quite similar manually (same handshape in different orientation), this exclusion affords a clearer picture of the relationship due to manual similarity alone.

The present results are not consistent with the production similarity data obtained by Locke (1970). Locke found the following pairs of handshapes to be rated as the most similar kinesthetically: K-P, B-Y, F-B, R-P, T-V, and X-K. Of these pairs, our subjects judged only the pair K-P to be highly similar. The pair F-B was judged to be only moderately similar. It is likely that the limited set of nine letters used by Locke, combined with a forced choice

Table 3

Correlations of STM Confusion Matrices with Manual Similarity Matrices

	Manual Similarity Matrix		
STM Confusion Matrix	Visual	Production	Combined
Conrad & Rush (1965)	-.17	.06	-.08
Wallace & Corballis (1973)	.45**	.51**	.50**

(** significant at .01 level or better)

Table 4

Correlations of STM Confusion Matrices (from Wallace & Corballis, 1973) with Combined Manual Similiarity Matrix

	Letter Set	
STM Confusion Matrix	Including G-Q	Excluding G-Q
Manually Trained Subjects		
List Length 4	.38**	.07
List Length 5	.48**	.45**
Orally Trained Subjects		
List Length 4	.29*	-.06
List Length 5	.37*	.20*

(* significant at .05 level or better, ** at .01 level or better) methodology, imposed a set of similarity relationships unrepresentative of the larger set of handshapes. It is also possible that subjects misinterpreted his instructions. Consider, for example, that the letter pair T-V was rated highly similar by Locke's subjects (in contrast to Weyer's study and the present one, which are the only other studies to include both the T and V handshapes). This letter combination is frequently produced by deaf individuals (in referring to television) and is quite easy to produce as a rapid sequence. If such an "ease of co-production" criterion was adopted by Locke's subjects, there would be little reason to expect our results to be similar.

In summary, our characterization of handshape similarity appears reasonably stable across both judgment modality and degree of experience. It is consistent with previous work in perceptual confusability, and is related in straightforward ways to patterns of interletter confusion in short-term memory. Future experiments can draw on these results either to manipulate systematically or to detect the use of manual codes in the processing of verbal stimuli.

References

Bellugi, U., Klima, E. S., & Siple, P. (1975). Remembering in signs. Cognition, 3, 93-125.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 55, 75-84.

Conrad, R. (1979). The deaf schoolchild. London: Harper & Row.

Conrad, R., & Rush, M. L. (1965). On the nature of short-term encoding by the deaf. Journal of Speech and Hearing Disorders, 30, 336-343.

Dodd, B., & Hermelin, B. (1977). Phonological coding by the prelinguistically deaf. Perception & Psychophysics, 21, 413-417.

Furth, H. G. (1973). Deafness and learning. Belmont, CA: Wadsworth.

Hanson, V. L. (1982a). Short-term recall by deaf signers of American Sign Language: Implications for order recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 572-583.

Hanson, V. L. (1982b). Use of orthographic structure by deaf adults: Recognition of fingerspelled words. Applied Psycholinguistics, 3, 343-356.

Hanson, V. L., Liberman, I. Y., & Shankweiler, D. (1984). Linguistic coding by deaf children in relation to beginning reading success. Journal of Experimental Child Psychology, 37, 378-393.

Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32, 241-254.

Klima, E., & Bellugi, U. (1979). The signs of language. Cambridge, MA: Harvard University Press.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1-27.

Lane, H., Boyes-Braem, P., & Bellugi, U. (1976). Preliminaries to a distinctive feature analysis of handshapes in American Sign Language. Cognitive Psychology, 8, 263-289.

Locke, J. L., (1970). Short-term memory encoding strategies of the deaf. Psychonomic Science, 18, 233-234.

Locke, J. L., & Locke, V. L. (1971). Deaf children's phonetic, visual, and dactylic coding in a grapheme recall task. Journal of Experimental Psychology, 89, 142-146.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352.

Neville, H. J., Kutas, M., & Schmidt, A. (1982). Event-related potential studies of cerebral specialization during reading. Brain and Language, 16, 316-337.

Quinn, L. (1981). Reading skills of hearing and congenitally deaf children. Journal of Experimental Child Psychology, 32, 139-161.

Reich, P. A. (1974). Visible distinctive features. In A. Makkai & V. B. Makkai (Eds.), The First LACUS Forum (pp. 348-356). Columbia, SC: Hornbeam Press.

Shepard, R. N. (1962). Analysis of proximities: Multidimensional scaling with an unknown distance function. Psychometrika, 27, 125-140, 219-246.

Stokoe, W. C., Casterline, D. C., & Croneberg, C. G. (1965). A dictionary of American Sign Language. Washington, DC: Gallaudet College Press.

Stungis, J. (1981). Identification and discrimination of handshape in American Sign Language. Perception & Psychophysics, 29, 261-276.

Treiman, R., & Hirsh-Pasek, K. (1983). Silent reading: Insights from congenitally deaf readers. Cognitive Psychology, 15, 39-65.

Wallace, G., & Corballis, M. C. (1973). Short-term memory and coding strategies in the deaf. Journal of Experimental Psychology, 99, 334-348.

Weyer, S. A. (1973). Fingerspelling by computer (Tech. Rep. No. 212). Stanford, CA: Stanford University, Institute for Mathematical Studies in the Social Sciences.

Zakia, R. D. & Haber, R. N. (1971). Sequential letter and word recognition in deaf and hearing subjects. Perception & Psychophysics, 9, 110-114.

Footnotes

[1] In skilled fingerspelling, letters of words are neither produced nor recognized as isolated letters. Rather, one finds evidence for coarticulatory effects in production (Reich, 1974) and facilitation of recognition in familiar clusters (Hanson, 1982b; Zakia & Haber 1971).

[2] The subjects in Weyer's experiment were 12 hearing subjects and 3 deaf subjects. Since the data of the deaf and hearing subjects were not presented separately, we do not know to what extent the overall characterization is representative of the deaf users of the language system.

[3] The study by Conrad and Rush (1965) used only 9 different letters: B, F, K, P, R, T, V, X, and Y. The study by Wallace and Corballis (1973) used only 10 letters: A, B, D, E, G, H, N, Q, R, and T. If we look at the visual and production similarity judgments obtained in the present study, it can be seen that the letters used by Conrad and Rush are relatively low in rated similarity (with the exception of K and P, which are moderately similar). The letters used by Wallace and Corballis have several pairs that were found by our techniques to be manually similar (namely, A-E, A-N, A-T, N-T, and G-Q).

SHORT-TERM MEMORY FOR PRINTED ENGLISH WORDS BY CONGENITALLY DEAF SIGNERS: EVIDENCE OF SIGN-BASED CODING RECONSIDERED

Vicki L. Hanson and Edward H. Lichtenstein†

Abstract. Shand (1982) found that deaf signers' recall of lists of printed English words was poorer when the American Sign Language translations of those words were structurally similar than when they were structurally unrelated. He presented these results as evidence of sign-based coding of printed words. This conclusion is challenged by the present finding that a group of hearing subjects, who were tested on Shand's stimuli and were unfamiliar with sign language, showed similar performance decrements on the lists of words having structurally similar signs. Alternative accounts of these findings for both hearing and deaf subjects are discussed.

The nature of short-term memory coding of printed words by deaf individuals is of considerable importance, both for theoretical and practical reasons. Theoretically, investigations in this area can provide insight into the role of speech coding in short-term ordered recall. Does the use of a speech code by hearing individuals derive from their experience with speech as a primary means of communication (Shand, 1982; Shand & Klima, 1981)? Or does speech coding provide an effective way of storing ordered information due to the highly sequential character of spoken language (Baddeley, 1979; Crowder, 1978; Healy, 1975)? If due to the primary use of a spoken language to communicate, then a language code rooted in another modality (e.g., a code based on visual/gestural sign language of deaf individuals) should be as effective a code for short-term ordered recall as a speech code. If, however, speech is an effective code for ordered recall due to its sequential properties, then deaf signers (who, in general, have received some speech instruction) may not be inclined to recode into signs, owing to the fact that signs involve simultaneous structuring of linguistic elements to a much greater extent than does speech (Klima & Bellugi, 1979).

In addition to these concerns, practical educational issues call for research on short-term memory coding by deaf individuals. In particular, research in this area addresses issues of coding in reading, such as whether deaf children can use speech coding in reading and whether sign coding can be used as an effective alternative to speech coding for deaf children in the

---

†National Technical Institute of the Deaf

Acknowledgment. This research was supported by Grant NS-18010 from the National Institute of Neurological and Communicative Disorders and Stroke and by Grant HD-01994 from the National Institute of Child Health and Human Development. We wish to thank Beth Schwenzfeier for her help in testing subjects, and Rena Krakow, John Richards, and Bruno Repp for their valuable comments on an earlier draft of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-85 (1986)]

acquisition of reading (see, for example, Conrad, 1979; Hanson, Liberman, & Shankweiler, 1984).

The presentation of American Sign Language (ASL) signs to deaf subjects for ordered recall has consistently provided evidence that information is coded in terms of the formational parameters (cheremes) of the signs; thus, presenting lists of formationally similar signs has lead to decrements in serial recall (Hanson, 1982; Poizner, Bellugi, & Tweney, 1981; Shand, 1980, 1982), and presenting lists of unrelated signs has lead to intrusion errors in which the incorrect items are formationally similar to the original signs (Bellugi, Klima, & Siple, 1975; Krakow & Hanson, 1985). In contrast, less consistent outcomes have been reported in experiments in which printed words have been presented to deaf subjects. Shand (1980, 1982) reported decrements in recall of printed words whose corresponding signs were formationally similar to other signs within the same list. However, the use of similar procedures by other researchers failed to obtain this same outcome, even when testing native signers of ASL (Hanson, 1982; Lichtenstein, in press). Furthermore, no evidence has been obtained of sign-based intrusions in deaf signers' recall of printed words (Krakow & Hanson, 1985). The failure to obtain evidence of sign-based coding with printed words cannot be attributed to insufficient power in the experimental design: Two of the studies that failed to find evidence of sign-based coding when printed words were presented found evidence of sign-based coding when *signs* were presented (Hanson, 1982; Krakow & Hanson, 1985).

The present paper focuses on the work of Shand (1980, 1982) in an attempt to resolve the discrepancy between his results and those reported by other investigators. A resolution of this issue is highly desirable given the importance of such findings for theories of short-term memory and their pedagogical implications.

Shand's procedure (following Baddeley, 1966) involved the use of experimental sets of words chosen to be similar along a given dimension. For this purpose, Shand had a phonetically similar set of words (SHOE, THROUGH, NEW, SHOW, NO, SEE, THREE, SEW) and a cheremically (sign) similar set of words (CANDY, ONION, APPLE, JAPANESE, JEALOUS, CHINESE, SOUR, BORED). The signs corresponding to each of the words in the cheremically similar set are similarly formed. Accuracy on lists of words taken from the experimental sets is compared, in this procedure, with accuracy on lists of words taken from a control set. As controls, Shand used four words from the phonetically similar set and four words from the cheremically similar set. The resulting set of words (SHOE, THROUGH, APPLE, JAPANESE, NO, SEE, SOUR, BORED) allowed for a comparison of accuracy of specific words when they were presented in the experimental set vs. when they were presented in the control set. Thus, each of the words in the control set was matched with a word in one of the experimental sets.

The subjects in Shand's study were eight congenitally, profoundly deaf signers of ASL. Three were native signers of ASL; the other five had a minimum of seven years of signing experience. On each trial, the subjects saw five of the words from a word set and were asked for immediate written ordered recall of these words. Shand found, both with list scoring (percent lists recalled perfectly) and with item scoring (percent words correctly recalled in the correct position), that the deaf subjects had poorer recall of words with presentations from the cheremically similar set than from the control set.

This same finding was obtained when analyzing only those four words that were common to the cheremic and control sets; recall of these words was less accurate when they were presented in lists from the cheremically similar set than when presented in lists from the control set. In contrast, there was no significant difference in accuracy between performance on the phonetically similar lists and the control lists for either list or item scoring.

Shand concluded that these results provided evidence for sign-based coding of printed words. However, confounds within his stimulus sets lead us to question this conclusion. First, semantic associations occurred among the words in the cheremically similar set (e.g., CHINESE-JAPANESE, CANDY-APPLE-ONION), of the sort that have been found to produce decrements in short-term memory for hearing subjects (Baddeley, 1966). In addition, Lichtenstein (in press) noted that the cheremically similar words had more letters (21% more), more syllables (36% more), and less visual distinctiveness (in terms of the range of number of letters per word) than the control words.

Shand reported no data from hearing subjects on his task. However, he stated that pilot studies with hearing subjects revealed recall decrements for the phonetically similar lists relative to the control or cherological lists. He did not state whether or not the hearing pilot subjects demonstrated recall decrements for the cheremically similar lists relative to the control lists. In view of the potential stimulus confounds noted above and the implications of his results, such a control group is vital.

Reported here are the results of hearing subjects tested in the short-term memory task of Shand. We used the same stimuli and procedures, making only one modification in the experimental design: Due to the fact that hearing subjects, in most short-term memory tasks, are able to recall more items than deaf subjects, we increased the number of words presented on a trial from five to six in an attempt to keep our error percentages roughly comparable to those obtained with the deaf subjects tested by Shand.

Method

Stimuli. The stimuli were the three word sets used by Shand (1980, 1982), as given above.

Procedure. On each trial, subjects were presented with six words from one of the three sets. The words were serially presented, at a 1 s presentation rate, on a computer controlled CRT display. They were presented in uppercase letters. There were 16 trials of words from each set, with each word occurring an equal number of times in each serial position. Trials were blocked, such that subjects saw all 16 lists from one set of words before proceeding to a different set. The subjects were tested individually, and the order of set presentation was varied between subjects. During the testing on lists from a given set, the eight words of that set were displayed, on index cards, for the subjects. Each set was typed, in different orders, on two index cards; some of the subjects saw the first ordering of words, while other subjects saw the second.

The instructions, spoken by the experimenter, informed subjects that they were to watch each of the six words presented on a trial, and to write their responses when the signal (***) appeared at the end of the trial. They were told to write down the words in the serial position in which they occurred on answer sheets that had the serial positions numbered 1-6 for each trial. The experiment was self-paced allowing subjects to initiate each trial by a key press on the computer keyboard.

Subjects. The subjects were eight normally-hearing college students in the New Haven area. They were paid for their participation in this 45-min experiment. None reported any familiarity with signs.

Results

Shown in Table 1 are the mean percentages of correct recall for both list scoring and item scoring. For comparison purposes, the results of the deaf subjects tested by Shand are also given in Table 1. As can be seen from the Table, the magnitude of the difference in accuracy between the cheremically similar set and the control set for these hearing subjects is comparable to that of Shand's deaf subjects. With list scoring, the accuracy was 23% less for the hearing subjects and 18% less for the deaf subjects. With item scoring, the accuracy was 8% less for the hearing subjects and 9% less for the deaf subjects. Analyses on the percentage correctly recalled by the hearing subjects confirmed that this performance difference between the cheremically similar lists and the control lists was significant. Analyses of variance indicated significant main effects of condition for both list scoring $F(2,14) = 11.79$, $p < .01$, $MSe = 204.93$, and item scoring, $F(2,14) = 13.57$, $p < .01$, $MSe = 34.13$. Post hoc tests revealed a significant difference in accuracy between the cheremically similar and control lists for both scoring procedures (Newman-Keuls, $p < .05$).

Table 1

Percentage Accuracy in Recall of Lists of Printed English Words.

	Lists recalled perfectly (%)		Items recalled correctly in the correct position (%)	
Stimulus Set	Hearing	Deaf[a]	Hearing	Deaf[a]
Control	77	62	91	86
Cheremically similar	54*	44*	83*	77*
Phonetically similar	43*	59	76*	84

Note: [a]from Shand (1982); * significantly different from control

The data of the hearing subjects differed from those of the deaf subjects in terms of the relative accuracy on the phonetically similar set. Consistent with Shand's statement regarding his hearing pilot subjects, the hearing subjects in the present study had difficulty in the recall of words from the phonetically similar sets. Post hoc tests revealed a significant difference in accuracy between the phonetically similar and control lists with both list scoring and item scoring (Newman-Keuls, $p < .05$). These hearing subjects, also consistent with Shand's observation regarding his hearing pilot subjects, were less accurate on the lists from the phonetically similar set than on lists from the cheremically similar set. This difference was statistically significant with item scoring (Newman-Keuls, $p < .05$), but not with list scoring (Newman-Keuls, $p > .05$).

Shand reasoned that using four words each from the phonetically and cheremically similar sets as the words in the control set would allow these matched items to serve as their own control; the ability to recall these matched items could be compared when they occurred in lists from the experimental set vs. when they occurred in lists from the control set, allowing for a determination as to the relative ability to recall particular words as a function of list type. The percentages of items correctly recalled on the four matched words in the control and experimental sets are given in Table 2. The error pattern on these matched words indicated recall decrements on both the cheremically similar lists, $t(7) = 4.25$, $p < .01$, and the phonetically similar lists, $t(7) = 2.72$, $p < .03$ (both tests two-tailed). Thus, for these subsets of the stimuli, as for the full set of stimuli, recall was less accurate when words occurred in experimental lists than when they occurred in control lists.

Table 2

Percentage Accuracy in Recall of Items that Appeared in Both the Control Set and an Experiment Set.

	Experimental Set			
	Phonetically similar (%)		Cheremically similar (%)	
Stimulus Set	Hearing	Deaf[a]	Hearing	Deaf[a]
Control	90	86	92	86
Experimental	77*	85	84*	78*

Note: [a]from Shand (1982); * significantly different from control

Discussion

The results reported here call into question Shand's (1982) conclusion that the deaf subjects' recall decrement on his cheremically similar lists of printed words can be taken as evidence of sign-based coding. When the same word lists were presented here to non-signing hearing subjects, their performance showed a decrement as well; a finding that, in this case, clearly cannot be attributed to sign-based coding. Rather, the greater semantic relatedness, word length, visual similarity, or number of syllables of the words in Shand's cheremically similar than control sets are likely to have led to the decrement for the hearing subjects. The same factor(s) may also have been responsible for the recall decrement for the deaf subjects.

However, the comparable decrement on the cheremically similar lists for both deaf and hearing subjects does not rule out the possibility of different underlying causes for these two subject groups. For example, it seems that Shand's deaf subjects would have been less likely than the present hearing subjects to have been influenced by the number of syllables in words. Although some deaf individuals, even native signers, have been found to use speech-based coding in short-term ordered recall (Conrad, 1979; Hanson, 1982; Lichtenstein, in press; Shand, 1980), these same studies have found that certainly not all deaf individuals do. The use of a speech code by prelingually, profoundly deaf persons appears to be related to a number of variables, including English proficiency and speech skill (Conrad, 1979; Hanson et al., 1984; Lichtenstein, in press). Shand's (1982) deaf subjects, as a group, showed no significant effect due to phonetic similarity; the present hearing subjects, however, did.

Although the possibility remains that the recall decrement for Shand's deaf subjects on the formationally (cheremically) similar lists was due to sign coding, the comparable results obtained here with hearing subjects clearly undercut Shand's argument. His conclusion must also be considered in light of the fact that his results are inconsistent with other studies in the literature in which deaf college students have served as subjects: Other studies using similar procedures (but different word sets) have shown performance decrements by deaf signers in the serial recall of formationally similar signs (Hanson, 1982; Poizner et al., 1981), but not in the serial recall of printed words having formationally similar signs (Hanson, 1982; Lichtenstein, in press). Moreover, although sign-based intrusion errors have been found in the serial recall of unrelated lists of signs (Bellugi et al., 1975; Krakow & Hanson, 1985), sign-based intrusions have not been found in the recall of lists of printed words (Krakow & Hanson, 1985). Taken together, these facts argue against Shand's conclusion that his results can be taken as evidence of sign-based coding of printed words.

References

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. Quarterly Journal of Experimental Psychology, 18, 362-365.

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), Processing of visible language (Vol. 1). New York: Plenum.

Bellugi, U., Klima, E. S., & Siple, P. (1975). Remembering in signs. Cognition, 3, 93-125.

Conrad, R. (1979). The deaf school child. London: Harper & Row.

Crowder, R. G. (1978). Language and memory. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic (pp. 331-376). Cambridge, MA: MIT Press.

Hanson, V. L. (1982). Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall. Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 572-583.

Hanson, V. L., Liberman, I. Y, & Shankweiler, D. (1984). Linguistic coding by deaf children in relation to beginning reading success. Journal of Experimental Child Psychology, 37, 378-393.

Healy, A. F. (1975). Coding of temporal-spatial patterns in short-term memory. Journal of Verbal Learning and Verbal Behavior, 14, 481-495.

Klima, E., & Bellugi, U. (1979). The signs of language. Cambridge, MA: Harvard University Press.

Krakow, R. A., & Hanson, V. L. (1985). Deaf signers and serial recall in the visual modality: Memory for signs, fingerspelling, and print. Memory & Cognition, 13, 265-272.

Lichtenstein, E. (in press). The relationships between reading processes and English skills of deaf college students: Part I. Applied Psycholinguistics.

Poizner, H., Bellugi, U., & Tweney, R. D. (1981). Processing of formational, semantic, and iconic information in American Sign Language. Journal of Experimental Psychology: Human Perception and Performance, 7, 1146-1159.

Shand, M. A. (1980). Short-term coding processes of congenitally deaf signers of ASL: Natural language considerations. (Doctoral dissertation, University of California, San Diego, 1980) Dissertation Abstracts International, 41.4, 1572A-1573A.

Shand, M. A. (1982). Sign-based short-term coding of American Sign Language signs and printed English words by congenitally deaf signers. Cognitive Psychology, 14, 1-12.

Shand, M. A., & Klima, E. S. (1981). Nonauditory suffix effects in congenitally deaf signers of American Sign Language. Journal of Experimental Psychology: Human Learning and Memory, 7, 464-474.

MORPHOPHONOLOGY AND LEXICAL ORGANIZATION IN DEAF READERS*

Vicki L. Hanson and Deborah Wilkenfeldt†

Abstract. Prelingually, profoundly deaf individuals, due to their hearing impairment, would not be expected to have the same access to phonological information as hearing individuals. They might therefore have difficulties in using phonological structure to relate different morphological forms of words. Deaf and hearing readers' sensitivity to the morphological structure of English words was tested in the present study by using a lexical decision (word/nonword classification) task. Target words were primed ten trials earlier by themselves (e.g., *think* primed by *think*), by morphologically related words (e.g., *think* primed by *thought*), or by orthographically related words (e.g, *think* primed by *thin*). Response times of both hearing and deaf college students to target words were facilitated when primed by themselves and also when primed by morphological relatives. Response times of subjects in neither group were facilitated to targets primed by orthographically related but morphologically unrelated words. These results indicate that deaf readers, like hearing readers, are sensitive to underlying morphophonological relationships among English words.

An appreciation of the morphological structure of English words has been demonstrated experimentally for hearing readers. In reading tasks, it has been shown that responses to words are facilitated by prior presentation of morphologically related words (Fowler, Napps, & Feldman, 1985; Murrell & Morton, 1974; Stanners, Neiser, Hernon, & Hall, 1979). Thus, for example, the word *walk* is more readily recognized when a morphologically related word such as *walks*, *walking*, or *walked* preceeds it than when no such morphological relative does. While facilitative effects due to priming by a semantic associate (e.g., *doctor* primed by *nurse*) generally appear not to persist beyond immediate testing (Dannenbring & Briand, 1982; Henderson, Wallis, & Knight, 1984), facilitative effects due to priming by a morphological relative have been found to persist for lags of at least 48 items (Fowler et al., 1985). Facilitation due to priming by a morphological relative is plausibly

---

**Language and Speech*, 1985, *28*, 269-280.

†Also University of Connecticut.

Acknowledgment. This research was supported by National Institute of Communicative Disorders and Stroke Grant NS-18010 and by National Institute of Child Development Grant HD-01994. We wish to thank persons at Gallaudet College who made it possible for us to conduct this research. We are also grateful to Laurie Feldman, Nancy Fishbein, Carol Fowler, and Rena Krakow for their comments on earlier versions of this manuscript and/or for their help in testing subjects.

attributed to a particular organization of the reader's internal lexicon in which morphologically related words are stored closely together (Fowler et al., 1985; MacKay, 1978; Stanners et al., 1979; Taft & Forster, 1975).

Not all morphologically related words have a common pronunciation and spelling of the shared morpheme (e.g., find-found), yet previous research suggests that hearing readers organize members of such disparate word "families" together in their lexicons (Fowler et al., 1985; cf. Stanners et al., 1979). This is apparently due to their ability, as speakers of English, to make use of the rules of the phonological component of the grammar, which relate underlying forms that are similar at the abstract level of morphophonological representation to forms that are different at the more concrete level of phonetic representation (see Chomsky & Halle, 1968).

The orthographic conventions of English appear to capture similarities at the morphophonological level and to exploit speakers' knowledge of the rules that render differences at the phonetic level. Thus, for example, the orthography represents the vowel in wide in the same way as the vowel in width (i.e., by the letter i) and represents the vowel in heal in the same way as the vowel in health (i.e., by the letters ea), reflecting the fact that at the abstract morphophonological level they are presumably the same. The speaker of English knows that in the first case the letter i represents the phone [ay] while in the second case it represents [I]. Note that not all orthographic similarities reflect morphophonological relationships (for example, cat is not related to catalyst). The evidence indicates that words are not stored closely together by virtue of their orthographic similarity alone in the mental lexicons of hearing readers (Feldman, in press; Murrell & Morton, 1974; Napps & Fowler, submitted).

It may not be surprising that hearing readers of English are able to make efficient use of an orthographic system that presupposes a knowledge of the underlying structure of the language. However, it can reasonably be asked whether deaf readers are able to make similarly efficient use of this orthographic system. Since deaf readers do not come to the task of learning to read English with the same experience with English phonology that hearing readers do, it is not clear whether they are able to take advantage of morphophonological relationships captured by the orthography. The present experiment investigates whether prelingually, profoundly deaf readers are able to acquire the knowledge of English phonology necessary to perceive the morphological relationships among written words that are observed in the orthography.

A technique that has been used to study morphological effects on lexical organization is that of repetition priming. This technique requires subjects to make a word/nonword response to each item during continuous presentation of letter strings. Lexical decision response times thus obtained are typically faster to the second presentation of a word than to the first (Forbach, Stanners, & Hochhaus, 1974), and are also typically faster to a word that has been preceded by a morphological relative (Stanners et al., 1979). This first type of facilitation will be referred to as identity priming; the second type as morphological priming. The present study uses the repetition priming technique to compare response times (RTs) to the first presentation of a word (e.g., think) with RTs to the same word when it has been preceded ten trials earlier by itself (think primed by think), by a morphologically related word (e.g., think primed by thought), and by an orthographically related word (e.g., think primed by thin). Although RT facilitation of the sort indicative of lexical effects have not been found when a word is preceded by an orthographically similar word (Napps & Fowler, submitted), it is possible that such facilitation will be obtained for deaf readers. In fact, this is exactly what would be expected if deaf readers fail to perform linguistic analyses of words, and instead, or in addition, organize words in their lexicons according to orthographic features.

Two types of morphological relatives are presented in this study: irregularly inflected forms (e.g., mouse-mice) and derived forms (e.g., prove-proof). Word pairs thus related differ phonetically and exhibit less orthographic overlap than do regularly inflected forms. As a result, readers can rely on neither phonetic nor orthographic similarity exclusively in recognizing the morphophonological relationships that hold between the members of each pair. Access to the underlying morphophonological representations would be necessary. For hearing readers, significant facilitation of words preceded by both irregularly related inflections and derivationally related words has previously been obtained (Fowler et al., 1985).

In the present experiment, the performance of deaf and hearing college students is compared. It should be borne in mind that the deaf college students who served as subjects represent the more advanced readers among the deaf population. These subjects were not tested in order to find out how deaf readers, in general, read, but rather to determine whether sensitivity to the underlying morphophonological relationships among words is possible at all in the presence of prelingual, profound deafness. A similar pattern of results for the hearing and deaf subjects would suggest that subjects in the two groups have a similar organization of their mental lexicons.

Method

Stimuli

Word triples were constructed consisting of a target word paired with both a morphological relative and an orthographically similar but morphologically unrelated word (e.g., think - thought - thin). The target words (e.g., think) and their orthographically similar primes (e.g., thin) always had at least the first three letters in common.

Preliminary lists of these triples were given to four deaf students from Gallaudet College, who were asked to indicate any words on the list that they did not know. Final stimulus lists were then constructed that excluded word triples from the preliminary list having one or more words that were judged to be unfamiliar.

The final list consisted of 24 word triples. For 14 of these triples, the morphological relative was an irregularly inflected form. For ten of the word triples, the morphological relative was a derivationally related form. The target words were generally high in frequency of occurrence in written English: 14 had a frequency of at least 100 per million words, 3 had a frequency of at least 50 per million, and the remaining 7 had a mean frequency of 27.6 per million (Thorndike & Lorge, 1944). A listing of the stimulus words is given in the Appendix.

Throughout the full experiment, each target word appeared once in each of three prime-target conditions. By appearing in each of these conditions, each target word served as its own control. The three prime-target conditions were (1) identity prime, in which the target word served as both target and prime (e.g., think being primed by think); (2) morphological prime, in which the target word was primed by a morphologically related word (e.g., think being primed by thought); and (3) orthographic prime, in which the target word was primed by an orthographically similar word (e.g., think being primed by thin). Although there was obviously some orthographic overlap between the target words and their morphological relatives, the orthographic overlap was less than that between the target words and their orthographic primes. The morphological primes had 2.13 letters in common with the target words (considering only common letters in the same word position); the orthographic primes had 3.42 letters in common with them. This difference was significant, $t(23) = 7.85$, $p < .001$.

Three experimental test lists were constructed so that in each list every target was tested in only one condition. Eight of the target words appeared in the identity prime condition, eight in the morphological prime condition, and eight in the orthographic prime condition. Each target followed the prime by a lag of ten items. In addition, there were twelve filler words per list.

To serve as an index of episodic (memory) effects, nonwords in the experiment were generated by replacing the initial consonant or consonant cluster of each word with another consonant or consonant cluster that made the letter string a nonword. For example, the nonword counterparts of the word triple less - least - lesson were dess - deast - desson. In list construction, the nonwords were treated similarly to their word counterparts, with each of the target nonwords preceded ten trials earlier by an identity prime, a morphological nonword prime, or an orthographic nonword prime. The final lists each contained 120 items, 60 of which were words and 60 nonwords.

A practice list of 30 items was constructed. The structure of the list was consistent with that of the experimental list.

Procedure

Stimulus presentation was controlled by a microcomputer. A trial began with the presentation of a warning signal (a "+") that appeared in the center of a CRT screen for 250 ms. The warning signal was then terminated and, following a 250 ms blank interval, a stimulus item was presented. Stimulus items were presented in uppercase letters in the center of the screen until the subject responded or until 5 seconds had elapsed. RT in milliseconds was measured from the onset of the letter string.

Subjects were instructed that they would be seeing strings of letters. They were told to indicate as rapidly and as accurately as possible whether or not each letter string was an actual English word by pressing one of two response buttons. If the letter string was a word, they were to press the YES response button. If the letter string was not a word, they were to press the NO response button. The YES button was pressed with the index finger of a subject's right hand, and the NO button with the index finger of the left hand. For the deaf subjects, the instructions were signed by a deaf experimenter who is a native signer of American Sign Language (ASL). For the hearing subjects, the instructions were spoken by a hearing experimenter. All subjects were individually tested.

Each subject saw all three experimental lists, list order being randomly drawn from the six possible orderings of the three lists. Thus, all subjects saw each target word once in each prime-target condition. Prior to testing with the three experimental lists, subjects were tested on the practice list.

Subjects

Deaf subjects were 14 students at Gallaudet College. All were prelingually and profoundly deaf with a hearing loss of 85 dB or greater (better ear average). All except one had deaf parents. The one subject who did not have deaf parents reported a family history of deafness (i.e., younger sibling, cousin). In all cases, then, the etiology of their hearing losses appears to have been hereditary deafness. The reading level of these subjects was assessed by means of the comprehension subtest of the Gates-MacGinitie Reading Tests (1978, Level F, Form 2), which was administered to each subject after completion of the experiment. The median reading grade equivalent of these subjects was 9.5 (Range: grade 3.3 to 12.9+).

Hearing subjects were 14 students at Yale University who reported no history of hearing impairment. The reading test was also given to these subjects, although in all but one case (a subject whose grade equivalent was 12.2), subjects' scores were so high as to be beyond the range for which the test was standardized (grade equivalent 12.9+).[1]

Results

Of interest here are RTs to words as targets in the three prime-target conditions compared with RTs to these same words as primes. Table 1 shows the means of the median RTs in the four conditions.

Table 1

Mean RTs (in ms) to Words as Primes and as Targets in
the Three Experimental Conditions.
The Mean Percentage errors are Given in Parentheses.

	Subject Group	
Condition	Deaf	Hearing
Prime	520 (7.4)	482 (5.4)
Target		
Identity Prime	494 (2.4)	458 (5.4)
Morphological Prime	505 (1.8)	467 (3.6)
Orthographic Prime	513 (5.1)	473 (4.2)

The median correct RTs were entered into analyses of variance on the within-subjects factor of condition (prime, target in the identity prime condition, target in the morphological prime condition, target in the orthographic prime condition) and the between-subjects factor of group (deaf, hearing). There was a significant main effect of condition in both the subjects, $F(3,78) = 8.12$, $p < .001$, and items analyses, $F(3,138) = 5.05$, $p < .005$, as well as when both were simultaneously considered, $F'\min(3,214) = 3.11$, $p < .05$. This effect did not significantly interact with group in either the subjects or the items analyses (both Fs < 1). Post hoc Tukey (hsd) tests indicated the source of this main effect: RTs to targets preceded by an identity or morphological prime were significantly faster than RTs to the same words as primes ($p < .05$), that is, RTs to targets primed by themselves and by morphological relatives were facilitated. There was no significant difference between RTs to targets in the identity prime and morphological prime conditions ($p > .05$). RTs to targets preceded by orthographic primes were not significantly facilitated ($p > .05$).

The error rates on the target words were low for both groups of subjects, as shown in Table 1. The analysis of the errors was generally consistent with the analysis of the RT data. An analysis of variance performed on the percentage of errors indicated no significant difference in error rates for the two groups in either the subjects or the items analyses (both Fs < 1). There was a main effect of condition in both the subjects, $F(3,78) = 5.69$, $p < .005$, and the items analyses, $F(3,138) = 3.95$, $p < .01$, which approached significance in the simultaneous consideration of both, $F'\min\ (3,211) = 2.33$, $.05 < p < .10$. Post hoc Tukey (hsd) tests indicated fewer errors to targets preceded by identity and morphological primes than to the same words as primes ($p < .05$). No other differences were statistically significant (all ps > .05). There was an interaction of condition X group in the subjects analysis, $F(3,78) = 3.19$, $p < .05$, but not in the items analysis, $F(3,138) = 2.22$, $p > .05$. The fact that this interaction was not significant in the items analysis suggests that the scores of just a few subjects deviated from the general pattern and that these deviant scores were responsible for the significant interaction in the subjects analysis. Inspection of the individual subjects' data supports this hypothesis. The deaf subjects generally produced fewer errors when the common morpheme had been previously accessed (i.e., in the identity and morphological priming conditions), while a few of the hearing subjects broke from this pattern, actually producing more errors to target words in the identity and morphological priming conditions than to priming words.

The nonword counterparts of the four conditions indicated that there was facilitation of nonwords primed by the identical nonword but not of nonwords primed by morphologically related or orthographically similar nonwords for either subject group. An analysis of variance on the RTs to the nonword conditions revealed a main effect of condition, $F(3,78) = 6.80$, $p < .001$, that did not interact with group, $F(3,78) = 1.97$, $p > .05$. There was no significant main effect of subject group, $F(1,26) = 3.09$, $p > .05$. The analysis of the error data indicated no significant main effects of either group, $F < 1$, or condition, $F(3,78) = 1.69$, $p > .05$, and no significant interaction of the two variables, $F(3,78) = 1.29$, $p > .05$. The means of the median RTs (and the mean percentage errors), collapsed across subject group, were 575 ms (9.4%), 550 ms (6.3%), 561 ms (9.4%), and 563 msec (7.2%), respectively, for the prime, the target in the identity prime condition, the target in the morphological prime condition, and the target in the orthographic prime condition. Post hoc Tukey (hsd) tests on the RTs indicated that the main effect of condition was due to faster RTs to nonwords when they served as targets in the identity prime condition than when they served as primes ($p < .05$). There was no significant facilitation of nonword targets in the orthographic prime condition ($p > .05$). Importantly, there also was no significant facilitation of nonword targets in the morphological prime condition ($p > .05$), suggesting that the facilitation obtained with words in the morphological prime condition was due to lexical, not episodic, effects. Moreover, with words in the morphological prime condition there were fewer errors to targets than primes, while, by contrast, in the nonword error data the percentage of errors did not significantly vary as a function of condition.

Because there is evidence that highly successful hearing readers/spellers are more sensitive to morphophonological relationships than are average or poor readers/spellers (Fischer, Shankweiler, & Liberman, 1985; Freyd & Baron, 1982), the question of whether deaf readers' sensitivity to morphophonological relationships varies as a function of reading proficiency was also examined. Correlations between deaf subjects' degree of priming in each of the target conditions and their grade level reading achievement were consistent with the notion that the better readers were more sensitive to morphological relationships than the poorer readers. Correlations were computed between scores on the comprehension subtest of the Gates-MacGinitie Reading Tests (1978) and their RT facilitation in the three target conditions. The measure of facilitation was the RT to primes minus the RT to targets. The correlations with facilitation in the identity prime condition ($r = .39$) and the morphological prime condition ($r = .41$) approached significance (both df=12, $.05 < p < .10$, one-tailed). There was no significant correlation between reading achievement and amount of facilitation in the orthographic prime condition ($r = .09$).

Discussion

The results of this experiment indicate that despite prelingual and profound hearing impairment, it is possible to acquire a sensitivity to the morphophonological structure of English words, even when morphological relations are expressed by orthographically dissimilar representations. In this experiment, deaf subjects, like hearing subjects, were facilitated in their response times to words that had been preceded by a morphological relative. Neither hearing nor deaf subjects were facilitated in their response times to words that had been preceded by an orthographically similar, yet morphologically unrelated, word.

Several pieces of evidence from the present study suggest that the obtained facilitation to target words in the morphological prime condition reflected lexical, not episodic influences. Episodic effects could arise in an experiment such as the present one because subjects remember seeing or responding to a particular letter string previously in the experiment. One indication of episodic effects in a repetition priming task is the presence of facilitation on nonwords (Feustel, Shiffrin, & Salasoo, 1983). There was, however, no such facilitation to nonword targets in the morphological prime condition, suggesting that the facilitation to target words in this condition can be attributed to lexical effects. Moreover, the fact that there was no facilitation to target word RTs in the orthographic prime condition is consistent with this interpretation. The number of common letters was greater in the orthographic than the morphological prime condition, and this greater overlap should lead to larger episodic effects; yet, there was no significant facilitation in the orthographic prime condition. The observed facilitation due to inflectional and derivational relationships is therefore consistent with conceptualizations of lexical organization in which morphologically related words are tightly associated (Fowler et al., 1985; Stanners et al., 1979; Taft & Forster, 1975).

Although the facilitation due to morphological priming did not differ significantly in magnitude from that due to identity priming, a look at Table 1 indicates that the facilitation is numerically somewhat smaller in the morphological prime condition. This greater facilitation may have been due, at least in part, to episodic influences acting in conjunction with lexical effects to facilitate RTs to targets in the identity prime condition (Feustel et al., 1983; Forster & Davis, 1984; Fowler et al., 1985). Evidence supporting this interpretation was obtained in the nonword data where significant RT facilitation occurred to target nonwords in the identity prime condition. Such an episodic influence may therefore also have affected response times to word targets in this condition.

The outcome of this study suggests that deaf readers, whose knowledge of English phonology may not be the same in all respects as that of hearing users of the language, do possess and utilize a knowledge of phonology that serves them well, at least in certain linguistic situations. Studies by other investigators have shown that deaf readers (also college students) are able to segment morphologically complex words into their stems and affixes and are aware that morphologically related words are semantically related (Hirsh-Pasek & Freyd, 1983, 1984; Lichtenstein, in press). The present study extends such findings by indicating that deaf readers' lexical organization is affected by the morphological composition of words.

Examination of the response time data in the present study reveals a somewhat smaller magnitude of identity and morphological priming facilitation than in previous studies (Fowler et al., 1985; Stanners et al., 1979). Procedural differences between the present study and earlier ones may account for this difference. In the present study, each target word occurred three times as a target item, once in each of the three experimental lists. Each subject was tested on all three lists. Studies have indicated that effects of identity and morphological priming may be apparent over relatively long time periods (Fowler et al., 1985; Scarborough, Cortese, & Scarborough, 1977). With respect to the present experiment, this suggests that response times to target words in the second and third lists tested would be facilitated not only by the priming word on that list, but also by prior presentations of the same and related words on earlier lists. This does not in any way invalidate the results of the present experiment; indeed, the results for the hearing subjects are quite consistent with other studies of hearing readers (Fowler et al., 1985; Stanners et al., 1979). However, the procedure used here would tend to diminish the magnitude of the facilitation effect since the response times to target words were averaged over the three lists.[2] Moreover, it is known that the magnitude of facilitation is greater for infrequently occurring words than for words that occur frequently (Forster & Davis, 1984; Scarborough et al., 1977). Nearly all the target words of the present study have a very high frequency of occurence in written English, a result of the necessity to obtain words within the vocabulary of all the subjects. This, too, is likely to have had the effect of diminishing the magnitude of any identity or morphological priming compared with other studies in the literature in which less frequently occurring words were used. In any case, even though the effects reported here are numerically somewhat smaller than those reported in previous studies, they are still statistically significant, in spite of the presence of factors that might have masked a larger priming effect.

As in experiments with hearing subjects both here and earlier (Napps & Fowler, submitted), the deaf subjects were not significantly facilitated in their response times to targets following an orthographically related prime. The implication of this result is that words are not, by virtue of orthographic (i.e., visual) similarity alone, closely associated in readers' lexicons. This is apparently the case for deaf as well as hearing readers of English. Although most morphologically related words do overlap orthographically a great deal, the present results suggest that formal similarity alone is not sufficient for organizing words together. Rather, what is required is a morphological relationship.

There was some indication in the present study that the degree of deaf readers' sensitivity to morphophonological relationships was related to their reading ability; specifically, the better readers were more sensitive to this level of linguistic structure than were the poorer readers. This finding is consistent with results reported for hearing subjects (Feldman, 1984; Freyd & Baron, 1982). Freyd and Baron (1982), for example, found that superior fifth-grade readers outperformed average eighth-grade readers in their ability to decompose morphologically complex words. Thus, the superior readers were better able to use the principles of English morphology. Based on such findings, it has been argued that skill in using the English orthography encompasses an ability to apprehend the morphological structure of words (Fischer, et al., 1985; Freyd & Baron, 1982).

In conclusion, it should be noted that this is not the only study in which evidence has been obtained that deaf readers are able to acquire some appreciation of the phonological component of English (see Dodd, 1980; Dodd & Hermelin, 1977; Hanson, Shankweiler, & Fischer, 1983). It does, however, extend previous work in finding that the organization of the mental lexicons of deaf readers is affected by morphological relationships captured at an abstract level by the phonological component of the grammar.

Such findings raise the question of the development of morphophonological sensitivity in prelingually, profoundly deaf readers. These readers have available to them some knowledge of the spoken language that they have acquired through experience with speaking and also lipreading. In addition, they have knowledge of word structure that has been acquired through reading and fingerspelling. (Fingerspelling is a manual representation of the orthography.) Each of these factors may contribute in part to the development of morphophonological sensitivity in prelingually, profoundly deaf readers (for further discussion, see Hanson, 1986; or Hanson et al., 1983). But it is likely that none of these factors can, by itself, account for the degree of morphophonological sensitivity observed in this experiment: Knowledge of the English sound system obtained without reference to its phonetic aspect is necessarily incomplete, and the morphology of English is represented by the orthography in a way that assumes prior familiarity with phonology on the part of the reader. The relative roles of the above factors in the development of the morphophonological sensitivity observed in this experiment remain to be determined, along with the nature of the contribution made by the innate linguistic abilities of deaf readers.

References

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Conrad, R. (1979). *The deaf school child*. London: Harper & Row.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology*, *36*, 435-444.

Dodd, B. (1980). The spelling abilities of profoundly pre-lingually deaf children. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 423-440). London: Academic Press.

Dodd, B., & Hermelin, B. (1977). Phonological coding by the prelinguistically deaf. *Perception & Psychophysics*, *21*, 413-417.

Feldman, L. B. (1984). *Structure of the noun system*. Paper presented at the meeting of the American Psychological Association, Toronto, Ontario.

Feldman, L. B. (in press). Phonological and morphological analysis by skilled readers of Serbo-Croatian. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production*. London: Academic Press.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, *112*, 309-546.

Fischer, F. W., Shankweiler, D., & Liberman, I. Y. (1985). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language*, *24*, 423-441.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, *2*, 337-539.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 680-698.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition*, *13*, 241-255.

Freyd, P., & Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior*, *21*, 282-295.

Gates-MacGinitie Reading Tests (Second Edition). (1978). Prepared by W. H. MacGinitie. Boston, MA: Houghton Mifflin Company.

Hanson, V. L. (1986). Access to spoken language and the acquisition of orthographic structure: Evidence from deaf readers. *Quarterly Journal of Experimental Psychology*, *38A*, 193-212.

Hanson, V. L., Shankweiler, D., & Fischer, F. W. (1983). Determinants of spelling ability in deaf and hearing adults: Access to linguistic structure. *Cognition*, *14*, 323-544.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language Processes* (pp. 211-226). London: Erlbaum.

Hirsh-Pasek, K., & Freyd, P. (August, 1983). *From print to meaning: Word identification through morphological analysis*. Paper presented at the meeting of the American Psychological Association, Los Angeles, CA.

Hirsh-Pasek, K., & Freyd, P. (October, 1984). *Taking the Latin and Greek out of English: Morphological analysis by hearing and deaf readers*. Paper presented at the meeting of the Boston University Conference on Language Development, Boston, MA.

Karchmer, M. A., Milone, M. N., Jr., & Wolk, S. (1979). Educational significance of hearing loss at three levels of severity. *American Annals of the Deaf*, *124*, 97-109.

Lichtenstein, E. H. (in press). The relationships between reading processes and English skills of deaf college students: Part II. *Applied Psycholinguistics*.

MacKay, D. G. (1978). Derivational rules and the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *17*, 61-71.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, 1974, *102*, 963-968.

Napps, S. E., & Fowler, C. A. (manuscript submitted for publication, 1984). *Formal relations among words and the organization of the mental lexicon*.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 1-17.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, *18*, 399-412.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*, 638-647.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College.

Footnotes

[1] Surveys in the United States and Canada have found that prelingually, profoundly deaf high school graduates generally only read with a grade equivalent of about third grade (Conrad, 1979; Karchmer, Milone, & Wolk, 1979). Therefore, the subjects of the present study were quite successful deaf readers, some of them being quite exceptional.

[2] One way to eliminate any effect due to presentation of target words in multiple lists would be to examine only the first list on which each subject was tested. However, within each list there were too few instances of each prime-target condition to produce reliable averaged response times. Thus, only the response times averaged over all three lists can be considered a reasonable measure.

APPENDIX

	Morphological Primes		
Target Words	Inflection	Derivation	Orthographic Primes
less	least		lesson
fight	fought		fig
more	most		moral
freeze	frozen		free
catch	caught		cat
rang	ring		ran
feet	foot		fee
teach	taught		tea
grind	ground		grin
find	found		fin
sunk	sink		sun
mouse	mice		mouth
tooth	teeth		too
think	thought		thin
speech		speak	speed
length		long	lend
voice		vocal	void
sale		sell	salad
sight		see	sigh
die		dead	diet
forty		four	fort
choice		choose	choir
singer		song	single
prove		proof	proverb

PERCEPTUAL CONSTRAINTS AND PHONOLOGICAL CHANGE: A STUDY OF NASAL VOWEL HEIGHT*

Patrice Speeter Beddor,† Rena Arens Krakow† and Louis M. Goldstein†

Abstract. To address the claim that listener misperceptions are a source of phonological shifts in nasal vowel height, the phonological, acoustic, and perceptual effects of nasalization on vowel height were examined. We show that the acoustic consequences of nasal coupling, while consistent with phonological patterns of nasal vowel raising and lowering, do not always influence perceived vowel height. The perceptual data suggest that nasalization affects perceived vowel height only when nasalization is phonetically inappropriate (e.g., excessive nasal coupling) or phonologically inappropriate (e.g., no conditioning environment in a language without distinctive nasal vowels). It is argued that these conditions, rather than the inherent inability of the listener to distinguish the spectral effects of velic and tongue body gestures, lead to perceptual misinterpretations and potentially to sound change.

Phonologists have long supposed that listener misperceptions are a source of phonological change (e.g., Durand, 1956; Jonasson, 1971; Ohala, 1981; Paul 1890/1970; Sweet, 1888). Listener misperceptions are presumably fostered by ambiguities in the acoustic signal with respect to articulation. That is, a given acoustic pattern may correspond (more or less closely) to more than one vocal tract configuration (e.g., [r] and [R] are spectrally similar, but articulatorily very different). If a language learner were to identify the articulatory source of an acoustic pattern incorrectly, (e.g., if [r] were perceived as [R]), then, in attempting to imitate that pattern the learner might produce the incorrectly reconstructed form rather than the original articulation. Thus, the similarity of certain segments in the acoustic domain could lead to their reinterpretation in the articulatory domain (e.g., [r] reproduced as [R]), and hence to sound change (e.g., /r/ > /R/ in German; Jonasson, 1971). (See Ohala, 1981 for further discussion.)

The present study addresses the claim that listener misperceptions are a source of phonological change within the domain of nasal vowel height. Phonologically, there is substantial synchronic and diachronic evidence of

---

*Phonology Yearbook, Vol. 3, in press.

†Also Yale University

Acknowledgment. This work was supported by NIH Grants HD-01994, HD-16591, and NS-07196. We thank Arthur Abramson, Bjorn Lindblom, and John Ohala for helpful comments on earlier drafts of this manuscript and Carol Fowler for stimulating and encouraging the experimental work described here.

[HASKINS LABORATORIES: Status Report on Speech Research SR-85 (1986)]

raising and lowering of nasal vowels in languages of the world. It has been suggested that shifts in nasal vowel height originate with the listener, who attributes some of the complex acoustic consequences of nasal coupling to changes in tongue height, thereby perceiving nasal vowels as higher or lower than their non-nasal counterparts (Chen, 1971; Ohala, 1974; Wright, 1980). As we will show below, this explanation for phonological shifts in vowel height is acoustically plausible, since some of the spectral consequences of coupling the nasal and oral tracts are similar to the effects of certain tongue body movements. However, this spectral similarity need not lead to perceptual confusion as to the articulatory source (i.e., tongue body versus velic gesture) of the spectral pattern. In fact, as we will show, nasal coupling does not affect perceived vowel height when nasalization of the vowel conforms to the phonetic and phonological structure of the listener's language. However, nasalization does influence perceived vowel height under certain conditions that are inconsistent with that structure, as when a conditioning environment for vowel nasality is absent or nasal coupling is excessive. It is argued that these conditions, rather than the inherent inability of the listener to distinguish the spectral effects of velic and tongue body gestures, lead to perceptual misidentifications and potentially to sound change.

Our goal, then, is to shed some light on the extent to which phonological shifts in nasal vowel height can be attributed to listener misperceptions. We therefore consider three types of data: phonological (section 1), acoustic (section 2), and perceptual (sections 3 and 4).

1. The Phonological Patterns

Diachronic and synchronic data from geographically distant and genetically unrelated languages indicate widespread phonological effects of nasalization on vowel height. For example, in French, synchronic morphophonemic alternations attest to historical lowering of high and mid vowels and raising of low vowels, as in (1) (where N represents any nasal consonant).

(1) French

[iN ~ æ̃]	e.g.,	fine/fin	'thin (fem/masc)'
[eN ~ æ̃]		plénitude/plein	'fullness/full'
[yN ~ œ̃]		une/un	'one (fem/masc)'
[ϕN ~ œ̃]		jeûne/(à) jeun	'fast/fasting'
[aN ~ ɑ̃]		planer/plan	'to glide/level'

Phonological studies comparing the height of contextual (allophonic) and non-contextual (phonemic or distinctive) nasal vowels to the height of corresponding oral vowels have found that, when differences occur, they are quite systematic across languages. Cross-language patterns of nasal vowel raising and lowering, based on Beddor (1983), are summarized in (2) (see Beddor (1983) for references). These patterns reflect synchronic allophonic and morphophonemic variation between oral and nasal vowel height in 75 languages, and are generally consistent with diachronic data and vowel inventory data from other cross-language surveys (Bhat, 1975; Foley, 1975; Ruhlen, 1978; Schourup, 1973).

(2) Cross-language patterns of nasal vowel raising and lowering

a. High (contextual and non-contextual) nasal vowels are lowered (e.g., nasalization lowers /i/ and /u/ in Bengali, Ewe, Gadsup, Inuit, and Swahili).
b. Low (contextual and non-contextual) nasal vowels are raised (e.g., nasalization raises /a/ in Breton, Haida, Nama, Seneca, and Zapotec).
c. Mid non-contextual nasal vowels are lowered (e.g., distinctive nasalization lowers /e/ and /o/ in Maithili, Portuguese, Shiriana, and Yuchi; distinctive nasalization lowers /e/ (but not /o/) in Hindi, Mixtec, and Kiowa Apache).
d. Mid back contextual nasal vowels are raised (e.g., /o/ or /ɔ/ is raised adjacent to N in Batak, Dutch, and Nama).
e. A mid front contextual nasal vowel is raised in a language where the corresponding back vowel is also raised (e.g., /e/ and /o/ are raised adjacent to N in Irish, Basque, and Havyaka Kannada); otherwise, mid front contextual nasal vowels lower are lowered (e.g., /e/ is lowered adjacent to N in Armenian, Campa, Fore, and Tewa, but /o/ does not shift in these languages).

These patterns show that the phonological effects of nasalization on vowel height involve the interaction of three factors: vowel height, vowel context, and vowel backness. Vowel height becomes centralized--that is, nasalization lowers high vowels and raises low vowels. Vowel context (presence or absence of an adjacent nasal consonant) affects mid vowel height, and distinguishes lowering of mid non-contextual nasal vowels from raising of mid contextual nasal vowels. Vowel backness also primarily affects mid vowels, but a front-back asymmetry holds for all vowels, such that front vowels are more likely to be lowered than back vowels. More specifically, lowering of a back nasal vowel in a language implies lowering of the corresponding front nasal vowel in that language (Beddor, 1983; see also Maddieson, 1984).

2. Acoustic Factors

The universality (in terms of genetic and geographic diversity) of these phonological patterns indicates that raising and lowering of nasal vowels are at least partially the result of phonetic constraints. Previously proposed phonetic explanations for shifts in nasal vowel height have appealed to articulatory (Lightner, 1970; Pandey, 1978; Pope, 1934; Straka, 1955), acoustic (Chen, 1971; Ohala, 1974; Wright, 1980), and perceptual (Haudricourt, 1947; Martinet, 1955; Ohala, 1983; Passy, 1890) constraints. Indeed, a comprehensive explanation of the phonological data may well need to recognize the interaction of several phonetic, as well as non-phonetic, factors. However, we will consider here but a single phonetic factor, the effect of nasalization on the first formant region of the vowel spectrum.

The main effect of vowel nasalization is in the vicinity of the first formant. According to acoustic theory of nasalization, coupling of the nasal tract to the oral tract adds a pole-zero pair to the low-frequency region of the vowel spectrum (Fant 1960; Fujimura & Lindqvist 1971; Stevens, Fant, & Hawkins, forthcoming). That is, the first formant F1 of the non-nasal vowel is replaced in the nasal vowel by a zero FZ and two formants, a shifted oral formant F1' and an extra nasal formant FN. FN is almost cancelled by FZ when coupling magnitude is small, but becomes more and more prominent as coupling increases. F1' typically differs in frequency, and has a wide bandwidth and low amplitude, relative to F1 of the uncoupled oral tract (Hawkins & Stevens, 1985; Stevens & House, 1956; Mrayati, 1975). Some of these spectral properties of nasal vowels are illustrated in Figure 1 by the vocal tract transfer functions for oral and nasal versions of a high front vowel generated on the Haskins Laboratories articulatory synthesizer (described below). As velopharyngeal coupling is increased from no coupling for oral [i] (top curve) to intermediate coupling (middle curve) and large coupling (bottom curve) for nasal [ĩ], the frequency of F1' shifts upwards and FN becomes increasingly prominent.

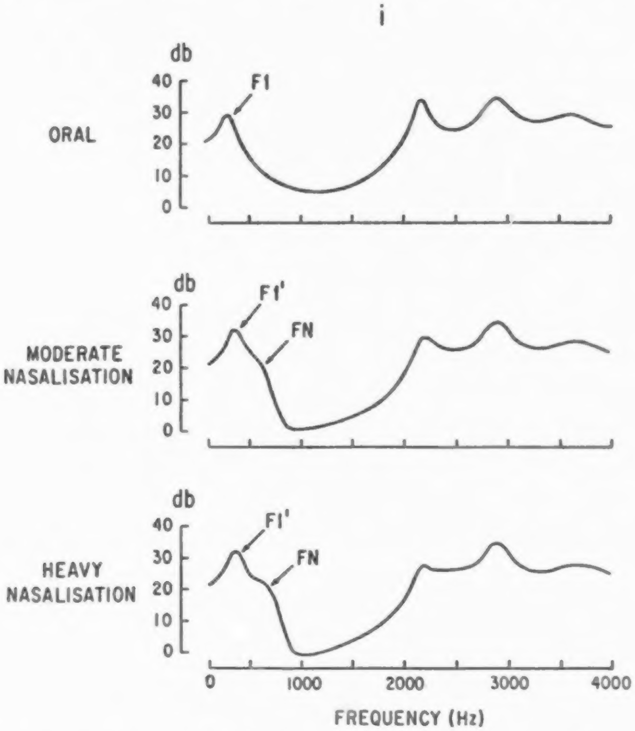


Figure 1. Vocal tract transfer functions for three versions of a high front vowel generated by articulatory synthesis: with no nasal coupling (top curve), with intermediate coupling (middle curve), and with large coupling (bottom curve). Nasal coupling shifted F1' upward relative to F1 and introduced FN, which showed increased spectral prominence with larger coupling.

2.1 First Formant Frequency

Shifts in the frequency of F1' in the nasal vowel relative to F1 in the oral vowel are of special interest here, since the frequency of F1 bears an inverse relation to vowel height. The acoustic theory of vowel nasalization predicts that nasal coupling increases the frequency of the first oral formant, that is, F1' > F1 (Fujimura & Lindqvist, 1971; Mrayati, 1975). This increase might lead us to expect nasalization to lower perceived vowel height. However, this expectation ignores the fact that the upward-shifted F1' is not necessarily the first peak in the nasal vowel spectrum. The lowest-frequency formant in the nasal vowel is located between F1 of the uncoupled oral tract and the lowest resonant frequency of the nasal tract when closed at the coupling end (probably 200-400 Hz; Fujimura & Lindqvist, 1971; Stevens et al., forthcoming). So when F1 of the oral vowel is relatively high (as in low vowels), the first formant of the coupled system is a low-frequency FN, as seen for low back [ɑ] and [ɑ̃] in Figure 2. In contrast, the first formant of the high nasal vowel that was shown in Figure 1 is the upwards-shifted low-frequency oral formant. It follows that the frequency of the first spectral peak is higher in a nasal vowel than in the corresponding oral vowel when the vowel is high, but lower when the vowel is low. This is consistent with the centralizing effect of nasalization on phonological vowel height discussed above and thus provides a tentative acoustic explanation for high nasal vowel lowering and low nasal vowel raising (Wright, 1980).

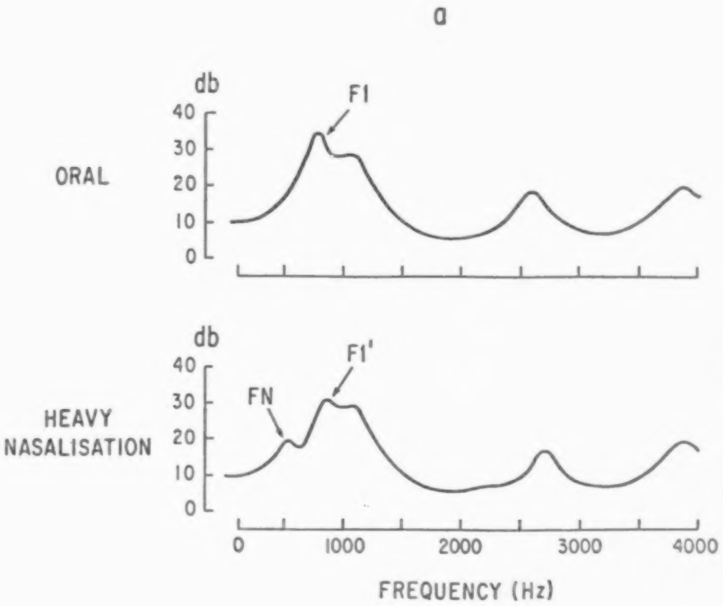


Figure 2. Transfer functions for oral [ɑ] (top curve) and nasal [ɑ̃] (bottom curve) generated by articulatory synthesis. Nasal coupling added a low-frequency FN and increased the frequency of F1' relative to F1.

We can use a model of acoustic-articulatory relationships to demonstrate how these acoustic factors could lead to a sound change. The ability of a listener (or language learner) to reproduce an arbitrary speech sound must depend on knowledge that links the acoustic properties to their articulatory origins. If such knowledge were always perfect, then there would be no sound changes (for this reason, in any case) at all. Thus, the knowledge brought to bear by the imitator is in some way imperfect (perhaps due to the inherent ambiguities mentioned earlier). As a model of an extreme case of such imperfection, let us imagine a listener (imitator) who has no knowledge of vowel nasalization at all and who reproduces any vowel as oral. How will such a listener reproduce nasal vowels?

This question can be answered using the equations developed by Ladefoged, Harshman, Goldstein, and Rice (1978) for calculating vocal tract shapes from formant frequencies. These equations are based entirely on oral vowels. Thus, the equations embody the acoustic-articulatory knowledge of a potential imitator ignorant of nasal vowels. We used these equations to calculate vocal tract shapes from the formants of the oral and heavily nasalized vowels shown in Figure 1. For the nasal vowel, the shifted oral formant (F1') was used as the lowest formant in the calculation. Figure 3(a) shows the vocal tract shape (of the articulatory synthesizer) that was actually used to generate the transfer functions in Figure 1 (except that the velar port was open in the nasal vowel). Figure 3(b) shows the recovered vocal tract shapes using the Ladefoged et al. equations. Ignoring obvious differences in the pharynx (the equations do not recover the shape of the lower pharynx), the recovered [i] is very much like the original. However, the shape recovered for [ĩ] is substantially lower. Thus, lack of knowledge of the effects of nasalization results in a high vowel being reproduced as a lower (oral) vowel. It is in this fashion that a sound change could develop. Of course, it is unlikely that any potential imitator has no knowledge of nasalization--the model simply shows the degree of lowering that would be expected in the most extreme case.

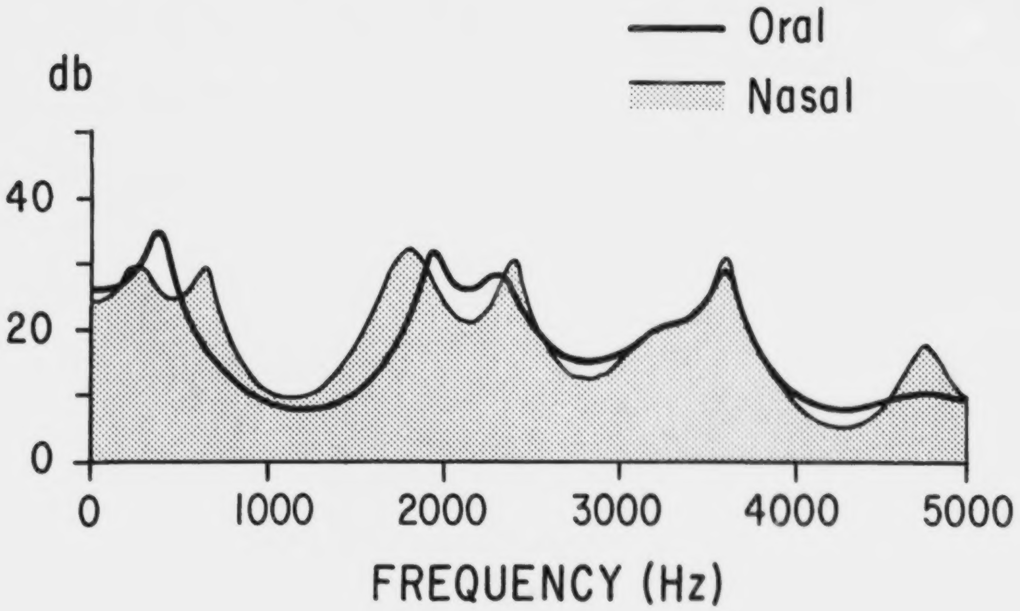
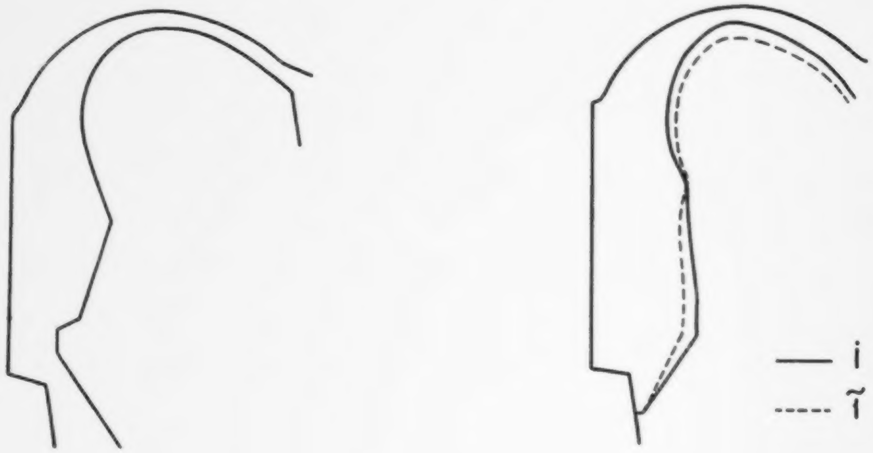


Figure 3. (a) Vocal tract shape of articulatory synthesizer used to compute the transfer functions in Figure 1. (b) Vocal tract shapes recovered from the formant frequencies of the transfer functions in Figure 1 (see text).

2.2 Center of Gravity

Although the effects of nasal coupling on the location of the first peak in the vowel spectrum are consistent with contraction of the height dimension, they do not appear to account for the front-back asymmetry in the phonological data. If we extend our acoustic measure of oral and nasal vowels to include not only frequency of the first spectral peak, but also frequency and relative amplitude of spectral peaks in the low-frequency region, we arrive at a more comprehensive explanation of the phonological patterns. Chistovich and her colleagues have found that perceived height of oral vowels reflects a "center of gravity" determined by the frequency and amplitude of spectral prominences in the F1-F2 region (Bedrov, Chistovich, & Sheikin, 1978; Chistovich & Lublinskaya, 1979; Chistovich, Sheikin, & Lublinskaya, 1979). Due to the complex acoustic effects of nasal coupling, nasalization can cause a shift in the center of gravity of the vowel spectrum that need not correspond to a parallel shift in the frequency of the first spectral peak. For example, in the naturally produced mid front vowels in Figure 4, the frequency of the first spectral peak is lower in nasal [ẽ] than in oral [e], but the overall effect of the pole-zero-pole combination in the low-frequency region of the nasal vowel is to pull up the center of gravity relative to the oral vowel.[1]



Original
(a)

Recovered
(b)

Figure 4. LPC spectra of oral [e] (unfilled) and nasal [ẽ] (filled) produced by a Hindi speaker. The nasal spectrum has a lower-frequency first peak, but a higher-frequency center of gravity, than the oral spectrum.

Beddor (1983) measured the center of gravity of oral and nasal vowel tokens from several languages by calculating the average frequency of the area under the spectral envelope in the F1-F2 region. This measure was consistently higher for [ĩ ẽ] than for [i e], lower for [æ̃ ɑ̃ õ] than for [æ ɑ o], and roughly the same for [ũ] and [u]. Assuming that an increase in center of gravity lowers perceived vowel height and a decrease raises perceived height, we would expect nasalization to perceptually lower /i e/, raise /æ ɑ o/, and have little effect on /u/. Thus, oral-nasal differences in center of gravity are consistent with the front-back asymmetry of the phonological data as well as high-low centralization.

3. Perceptual Validation

We have shown that the effects of nasal coupling on the low-frequency region of the vowel spectrum are generally consistent with the phonological patterns of nasal vowel raising and lowering. However, the acoustic data "explain" the phonological shifts only if the listener is misled by the resemblance between spectral changes due to nasal coupling and those due to tongue body movements; that is, if the listener has imperfect knowledge of acoustic-articulatory relations, as discussed above. Rather than assign all of the spectral consequences of nasalization to the velic gesture that couples the oral and nasal tracts, the listener must incorrectly attribute some of those spectral effects to a tongue gesture that modifies the oral tract configuration. Is there empirical evidence of such misperceptions? In answering this question, we hope to shed light not only on the role of listener misperceptions, but also on the relevance of context and speaker variability to vowel height shifts.

3.1 Perception of Non-Contextual Nasal Vowels

Several studies have investigated the perception of nasal vowel height. Wright (1980) produced natural oral and nasal vowels having the same tongue configuration, but differing in the position of the velum. All possible pairings of oral and nasal vowels were presented to listeners for similarity judgments. The perceptual vowel space constructed from listener responses showed centralization of nasal vowel height relative to oral vowel height. Acoustic analysis of the vowels indicated that this centralization did not always correlate with frequency differences in F1' versus F1, but might be partially due to the extra low-frequency FN in the nasal vowels.

In contrast to Wright's articulatorily matched vowels, Beddor (1984) paired oral and nasal vowels generated by formant synthesis. Listeners heard vowel sets in which a continuum of oral vowels (varying in the frequency of F1) was compared with a nasal vowel standard; they selected the oral vowel in each set that sounded most similar to the nasal standard. Listeners rarely chose the oral vowel in which F1 frequency was the same as F1' frequency in the nasal vowel. In general, listeners' choices were pulled towards FN of the nasal vowel: when FN frequency was low, the oral vowel chosen as the "best match" had a relatively low F1 frequency; when FN frequency was high, the oral match had a high F1. Apparently (as in Wright's study), shifts in the spectral center of gravity due to the added nasal formant affected perceived vowel quality.

In a recent study reported in Krakow, Beddor, Goldstein, and Fowler (in preparation), we used articulatory synthesis to investigate the effects of nasal coupling on perceived vowel height. The Haskins articulatory synthesizer allows specification of a mid-sagittal outline of the vocal tract by means of the positions of six articulatory parameters: jaw, hyoid, tongue body center, tongue tip, lips, and velum. The program computes the area functions for the specified vocal tract outlines. Speech output is obtained after acoustic transfer functions are computed for these area values (see Abramson, Nye, Henderson, & Marshall, 1981; Rubin, Baer, & Mermelstein, 1981).

In our study, we focused on the English /ɛ/-/æ/ contrast and generated seven vowels by systematically lowering and retracting the tongue body, as shown in Figure 5. These vowel shapes were then embedded in an articulatory context appropriate for [b_d] and two 7-step continua were generated: oral [bɛd-bæd] and nasal [bɛ̃d-bæ̃d]. The use of articulatory synthesis ensured that the only difference between the continua was that the velopharyngeal port was open during the vowel portion of the nasal, but not the oral, stimuli. Identification tapes for the two continua consisted of 10 tokens of each stimulus arranged in random order. Tapes were played to 12 phonetically naive native speakers of American English, who labeled the stimuli as *bed* or *bad*; they had no difficulty identifying the nasal vowel stimuli as such.

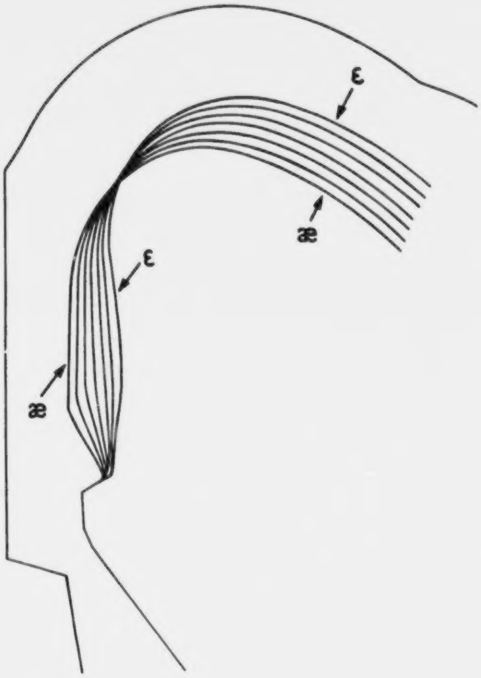


Figure 5. Vocal tract outlines of the seven steady-state vowel configurations specified by lowering and retracting the tongue body in equal articulatory steps from /ɛ/ to /æ/.

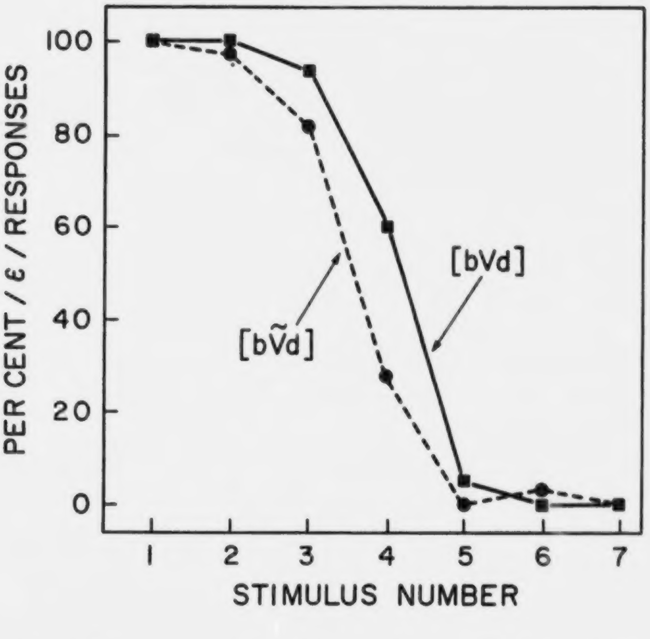


Figure 6. Pooled identifications functions (n = 12) for the oral [bɛd-bæd] (squares) and the nasal [bɛ̃d-bæ̃d] (circles) continua.

The identification functions in Figure 6 show the percent /ɛ/ responses for both the oral (indicated by the squares) and the nasal (the circles) stimuli. There were fewer /ɛ/, and therefore more /æ/, responses to the nasal vowels than to the oral vowels; that is, nasalization lowered perceived vowel height. This perceptual lowering is consistent with certain acoustic consequences of coupling the nasal tract to an /ɛ/-like oral tract configuration. For example, Figure 7 gives the transfer functions for stimulus 4, which listeners more often labeled /ɛ/ when oral but /æ/ when nasal. Although FN and F1' of the nasal vowel straddle F1 of the oral vowel, the predominant peak in the low frequencies of the nasal vowel spectrum is the upward-shifted F1'. The identification data can be interpreted as a tendency for listeners to associate the frequency shift induced by nasal coupling with lowering of the tongue body.

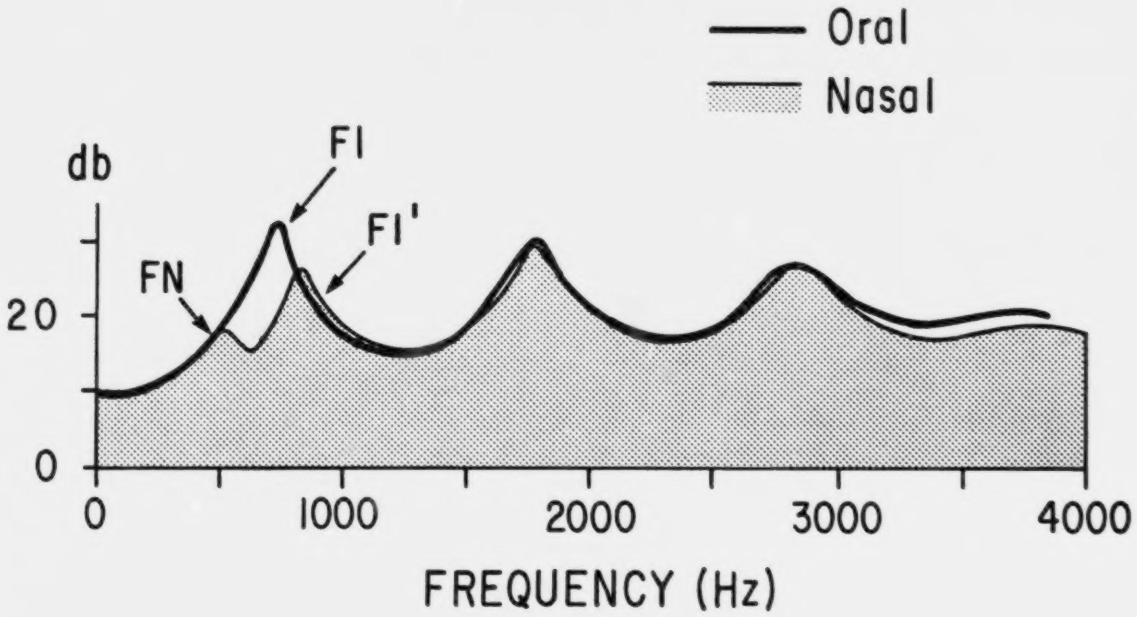


Figure 7. Transfer functions for the steady-state vowel portion of stimulus 4 from the oral (unfilled) and nasal (filled) continua. The first spectral peak has a lower frequency in the nasal vowel than in the oral vowel, but the predominant spectral peak in the nasal vowel is the upward-shifted F1'.

Our perceptual findings, like those of earlier studies, suggest that listeners have difficulty assessing the individual contributions of vowel quality and nasalization to the spectral shape of the nasal vowel. Listeners may have attributed the spectral shifts in part to nasalization, thus leading to the perception of a nasal vowel differing in height from the corresponding oral vowel. Alternatively, the spectral shifts may have been attributed entirely to oral tract shape, leading to the percept of a shifted oral vowel. Nonetheless, the data clearly show that spectral effects of nasalization on vowels produced in isolation or in an oral context (i.e., non-contextual vowel nasalization) are prone to misinterpretation by American listeners. And yet it would be premature to interpret these findings as evidence that listener misperceptions are a source of phonological shifts in nasal vowel height.

These listeners may have been prompted to resolve the spectral effects of non-contextual nasalization in terms of a tongue gesture only because of their unfamiliarity with distinctive nasal vowels.[2] That is, such misperceptions might not occur when the context for nasality is phonologically appropriate for the listener. To test this possibility, we need to look at the perception of non-contextual nasal vowels in languages with distinctive vowel nasalization and also at the perception of contextual nasal vowels (i.e., nasal vowels in the immediate context of a nasal consonant) in languages with anticipatory or perseverative nasalization. Some of our research addresses the second of these two issues.

3.2 Perception of Contextual Nasal Vowels

Lowering of the velum for a nasal consonant has been found to begin during a preceding vowel to some degree in all languages investigated. Substantial anticipatory vowel nasalization has been documented for many languages, including English (Ali, Gallagher, Goldstein, & Daniloff, 1971; Clumeck, 1976; Malécot, 1960; Moll, 1962). In Krakow et al. (in preparation), we tested our English-speaking subjects' perception of not only oral [bɛd-bæd] and non-contextual nasal [bɛ̃d-bæ̃d], but also contextual nasal [bɛ̃nd-bæ̃nd]. We speculated that in the [bṼnd] condition, the spectral effects of nasalization on the vowel might be attributed to an anticipatory velic lowering gesture for the nasal consonant, thus allowing more accurate assessment of vowel configuration than in the [bṼd] condition.

Support for our speculation is provided by previous studies in which listeners were shown to be sensitive to coarticulatory information. In a study of vowel nasality, Kawasaki (1986) reported that perceived nasality of vowels in [mVm] syllables was enhanced by attenuation of the adjacent nasal consonants. Her results suggest that listeners partially "factored out" vowel nasalization when the conditioning environment for nasalization was perceptually salient. Ohala, Kawasaki, Riordan, and Kaisse (in preparation; see also Ohala, 1981) looked at listeners' ability to recognize the coarticulatory fronting effects of apical consonants on adjacent /u/. They found that vowels ranging from [i] to [u] were more often labeled as back /u/ when flanked by apical consonants ([s_t]) than by labial consonants ([f_p]), that is, listeners apparently discounted some of the frontness of the vowels in the apical context as due to coarticulatory effects. Other studies have suggested that listeners are able to factor out coarticulatory effects not only of consonants on vowels, but also of vowels on consonants (e.g., Fowler, 1984; Kunisaki & Fujisaki, 1977; Mann & Repp, 1980; Whalen, 1981) and vowels on vowels (Fowler, 1981). These data all suggest that knowledge of how phonetic units are coproduced influences speech perception. (More specific theoretical accounts of such facts have been proposed in Fowler, 1983; Liberman & Mattingly, 1985.) We thought that such knowledge might enable listeners to distinguish the effects of nasalization from those of tongue shape on the spectrum of a contextual nasal vowel.

In our study, the contextual nasal condition [bɛ̃nd-bæ̃nd] was matched as closely as possible to the oral and non-contextual nasal conditions described above. All vowel stimuli had the tongue shapes shown in Figure 5. The contextual nasal continuum was the same as the non-contextual nasal continuum, except that the velopharyngeal port in the contextual nasal stimuli was open not only during the vowel, but remained open (at 16.8 mm^2) for 80 ms of the 137 ms alveolar occlusion, yielding natural-sounding [bṼnd] sequences. Since the steady-state portions of corresponding contextual and non-contextual nasal vowels were identical, we hypothesized that if the perceived height of nasal vowels were strictly a function of their spectral characteristics, then listeners would judge vowel height to be the same in the two nasal conditions. However, if tacit knowledge of anticipatory nasalization in English enabled listeners to factor out the spectral effects of contextual nasalization, then perceived height of the contextual nasal vowels would be more, if not exactly, like that of the oral vowels.

Labeling responses to the contextual nasal stimuli were obtained from the 12 subjects who identified the oral and non-contextual nasal vowels. The experimental procedure was the same as described above, except that subjects labeled the nasal stimuli as bend or band (as opposed to bed or bad). In Figure 8, the identification responses to the contextual nasal [bṼnd] stimuli (the diamonds) are compared with the [bVd] and [bṼd] functions from Figure 6. Notice that the point at which subjects shifted from /ɛ/ to /æ/ responses (i.e., the 50% crossover point) in the [bṼnd] condition was the same as in the [bVd] condition; that is, contextual nasalization had no effect on perceived vowel height.

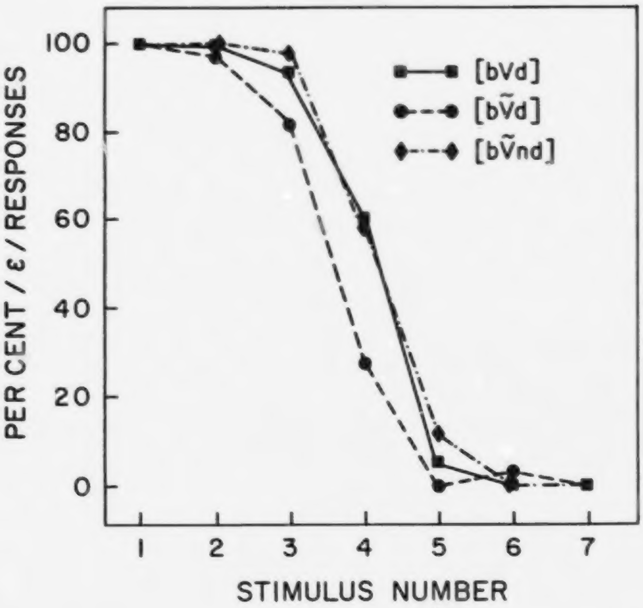


Figure 8. Pooled identification function (n = 12) for the contextual nasal [bɛ̃nd-bæ̃nd] continuum (diamonds) as compared with the oral [bɛd-bæd] and non-contextual nasal [bɛ̃d-bæ̃d] functions (see Figure 6).

3.3 Discussion

The perceptual data call into question simplistic accounts of the relation between listener misperceptions and nasal vowel height shifts. First, that listeners did not misjudge nasal vowel height when provided with a conditioning environment for vowel nasalization fails to support the idea that changes in contextual nasal vowel height are due to listeners misinterpreting the spectral effects of nasalization as cues for vowel height. Secondly, although the finding that listeners misjudged nasal vowel height in the absence of a conditioning environment might appear to support listener misinterpretations as a source of non-contextual height shifts, this finding must be evaluated in light of the language background of the listeners. Our explanation for perceptual lowering of the non-contextual nasal vowels is that our American listeners did not expect a nasal vowel in the context [b_d] and consequently perceived the spectral changes introduced with nasal coupling as due at least in part to tongue configuration. This reasoning prompts us to expect different results if we were to obtain judgments of the non-contextual nasal vowels from listeners whose native language has distinctive vowel nasalization. Since these listeners "expect" nasal vowels to occur in oral (as well as nasal) contexts, we hypothesize that non-contextual nasalization would have less of an effect--or perhaps no effect--on their perception of vowel height.

If listeners can separate the spectral effects of nasal coupling from those of tongue configuration, how then do we explain phonological raising and lowering of nasal vowels? We could, of course, turn to articulatory or even non-phonetic explanations, but the consistent correlations between the acoustic effects of nasalization and the phonological patterns make us reluctant to reject an acoustic-perceptual approach. We can maintain that listener misperceptions lead to shifts in nasal vowel height if we can show that normal perceptual processing occasionally fails. Specifically, since listeners normally distinguish the acoustic consequences of velic versus tongue body gestures, we need to show that this distinction can break down under certain conditions. In the next section, we consider two conditions that could lead to perceptual confusion and potentially to sound change.

4. Sources of Perceptual Confusion

4.1 Loss of Conditioning Environment

Ohala (1981, 1983) has argued that many sound changes in which loss of the conditioning environment co-occurs with the conditioned change can be explained by the listener's failure to detect the conditioning segment. We believe a similar argument provides a tentative explanation for shifts in non-contextual nasal vowel height.

In the vast majority of languages that have distinctive nasal vowels, such vowels evolved from earlier sequences of phonemic oral vowels followed by nasal consonants (Ferguson, 1963) or preceded by nasal consonants (Hyman, 1972). One account of phonemicization of vowel nasalization with concomitant nasal consonant loss is that the perceptual salience of vowel nasality increased as the perceptual salience of the conditioning nasal consonant decreased (see Kawasaki, forthcoming).[3] However, at the transition stage, distinctive vowel nasalization is not fully integrated into the language. If listeners do not expect non-contextual nasal vowels but also do not perceive the now-weakened nasal consonant, then they might attribute the acoustic effects of vowel nasalization to either (A) nasal coupling, (B) change in tongue configuration, or (C) both nasal coupling and change in tongue configuration. Under these conditions, we would expect /VN/ or /NV/ to result historically in (A) /Ṽ/ with nasalization but no height change, (B) /V'/ with height change but no nasalization, or (C) /Ṽ'/ with height change and nasalization.

Language data provide evidence of all three types of phonological change. There are numerous type A languages in which nasalization has no marked effect on vowel height. For example, nasal vowel inventories were reportedly the same as oral vowel inventories in 77 of the 155 languages with phonemic nasal vowels surveyed by Ruhlen (1978). Possible examples of type B change include Greek *en > a (*hekenton > hekaton; Foley, 1975), Colloquial Tamil final eːn > æː (Bhat, 1975) and Old Norse, in which i and u lowered when a following nasal consonant was lost, but the nasality of the lowered vowels is uncertain (Bhat, 1975). Type C languages are more difficult to identify, since distinctive nasalization and height shift must be shown to have occurred more or less simultaneously. One such language appears to be French. Accounts of the evolution of French low non-contextual nasal vowels from non-low vowels followed by nasal consonants disagree on the relative order of distinctive nasalization and vowel lowering (compare Entenman, 1977; Haden & Bell, 1964; Martinet, 1965; Pope, 1934), but the disagreement itself suggests that the two changes occupied roughly the same time period.

Evidence of type A languages (/VN/ > /Ṽ/) indicates that nasal consonant loss is not a sufficient condition for phonological shifts in nasal vowel height. At the same time, the existence of type B (/VN/ > /V'/) and type C (/VN/ > /Ṽ'/) languages suggests that nasal consonant loss is a possible trigger for such shifts. These phonological data correspond to our experimental results with American English speakers showing perceptual height shifts in [bṼd] sequences, although our results fail to distinguish whether listeners attributed all (as in type B languages) or only some (as in type C languages) of the spectral consequences of nasalization to tongue height.

Our claim, then, is that listeners' ability to distinguish the acoustic consequences of velic versus tongue body gestures might break down if the listener encounters a nasal vowel, but neither detects a conditioning nasal consonant nor expects non-contextual vowel nasalization. We hypothesize that these conditions lead to ambiguity as to the nasality of the vowel. This uncertainty could in turn lead to changes in vowel height if the listener were to resolve at least some of the acoustic effects of nasalization in terms of tongue configuration.

We have argued that, as a result of nasal consonant loss, there might be perceptual ambiguity leading to changes in vowel height. The next section postulates a second source of listener misperceptions that could influence the height of not only non-contextual, but also contextual, nasal vowels.

4.2 Variability in Production

Most of our discussion of nasal vowels has approached vowel nasalization as a binary distinction, such that vowels are either nasal or non-nasal. But there is considerable variation in degree of vowel nasalization across vowel tokens, types, and contexts, as well as across speakers and languages (Benguerel, Hirose, Sawashima, & Ushijima, 1977; Clumeck, 1976, Henderson, 1984; Ohala, 1971a; Ushijima & Sawashima, 1972). And as we have already seen (section 2), different magnitudes of nasal coupling have different effects on the vowel spectrum.

What influence, then, might variability in degree of nasalization have on vowel height? Consider, for example, a vowel followed by a nasal consonant. It seems reasonable to assume that the presence of the nasal consonant gives rise to certain expectations about the nasality of the vowel. If expectations are met, listeners should be able to factor the correct amount of vowel nasalization out of the vowel spectrum. But they might factor out too much (i.e., overcompensate) if nasalization is unexpectedly weak, or too little (undercompensate) if nasalization is excessive. (See Ohala, 1981, 1983 for discussion of the possible role of overcompensation in sound change.) Both errors could affect perceived vowel height: overcompensation would reverse the direction of the height shifts predicted by acoustic factors, while undercompensation would yield the predicted shift.

Some of our [bɛ̃nd-bæ̃nd] results address this issue. The data presented in section 4 were for a moderate (i.e., natural-sounding) amount of velopharyngeal port opening. But listeners were also tested on contextual and non-contextual nasal vowel stimuli produced with a small port opening (where nasalization was judged by the experimenters as perceptually weak) and a large port opening (where nasalization was judged as perceptually strong). The small port opening should raise the perceived height of the nasal vowels relative to the oral vowels if listeners overcompensate for weak nasalization; the large port opening should lower nasal vowel height if listeners undercompensate for strong nasalization.

Figure 9 gives the identification responses to the contextual [bṼnd] and non-contextual [bṼd] stimuli with small (7.2 mm²) and large (24.0 mm²) port

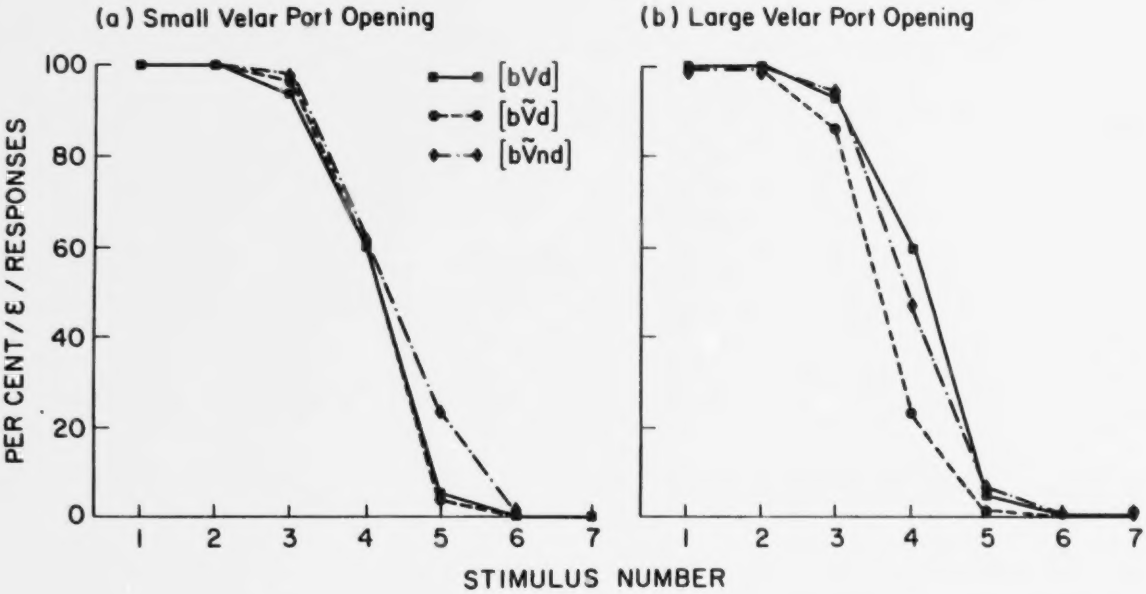


Figure 9. Pooled identification functions (n = 12) for the nasal continua generated with a small velar port opening (a) and a large port opening (b). The oral function is redrawn for comparison.

openings. In comparison with the oral [bVd] function, the nasal functions in Figure 9a show that, although weak non-contextual nasalization did not influence perceived vowel height, weak contextual nasalization slightly raised perceived height (i.e., there were more /ɛ/ responses to the [bṼnd] stimuli than to either the [bVd] or the [bṼd] stimuli). Since our American English listeners presumably expected nasalization in the contextual, but not the non-contextual, nasal conditions, our data suggest that listeners overcompensated for unexpectedly weak nasalization. This finding lends support to the speculation by Ohala (1983) that phonological raising of mid contextual nasal vowels (as opposed to lowering of mid non-contextual nasal vowels) might be explained by listener overcompensation for contextual nasalization.

In contrast, Figure 9b shows that strong non-contextual and contextual nasalization lowered percieved vowel height, with the non-contextual nasal vowels exhibiting greater lowering (i.e., the [bṼd] stimuli elicited the fewest /ɛ/ responses and the [bVd] stimuli the most). We interpret these results as evidence that listeners undercompensated for unexpectedly strong nasalization.

The implication of these findings for sound change is that variability in degree of vowel nasalization could cause perceptual uncertainty as to the relative contributions of the nasal and oral tracts to the vowel spectrum. Both weak and strong nasalization could lead to height shifts because of listener failure to correctly assess these contributions.

5. Further Questions and Conclusion

Several issues concerning nasal vowel height have not yet been resolved. We have not yet studied the perception of nasal vowel height by speakers of a language with distinctive vowel nasalization. While we have speculated that such listeners would show little or no effect of non-contextual nasalization on perceived vowel height, absence of these data clearly limits our understanding of listeners' ability to factor out the effects of nasal coupling on the vowel spectrum. Unfortunately, this experiment may prove to be difficult to do, since many languages with distinctive nasal vowels show vowel quality differences between oral and nasal vowels, or phonotactic constraints against /CVNC/ sequences, or both (severely limiting the use of our stimuli for these purposes).

Another concern is that the timing of the velic gesture, like its magnitude (i.e., size of velopharyngeal opening), can differ in speakers' productions of nasal vowels, depending on the quality of the vowel, the speaker, and the language (Clumeck, 1976). We still need to determine how temporal variability in the onset of the velic gesture affects the perceived height of nasal vowels. However, the present work leads us to conjecture that, for a given language, there is an "expected" temporal pattern and that deviations from that pattern (e.g., premature velic lowering) would lead to perceptual ambiguity and perhaps phonological change in nasal vowel height.

In summary, we have seen that there are consistent cross-language phonological patterns of nasal vowel height defined by the interaction of vowel height, context, and backness. We have also seen that a primary acoustic consequence of nasalization is the introduction of a pole-zero pair in the vicinity of F1, the effect of which is to shift the center of gravity in nasal vowel spectra relative to corresponding oral vowel spectra. These center of gravity shifts can account for two important variables in the phonological data, vowel height and vowel backness, and therefore provide phonetic motivation for most of the phonological patterns if these acoustic effects of vowel nasalization affect perceived vowel height. The perceptual data suggest that listeners misperceive nasal vowel height only when nasalization is phonetically inappropriate (e.g., excessive nasal coupling) or phonologically inappropriate (e.g., no conditioning environment in a language without distinctive nasal vowels). If inappropriate nasalization were unique to the laboratory setting, then these perceptual findings would oblige us to reject the claim that listener misperceptions are a source of nasal vowel height shifts in natural languages. However, even though inappropriate nasalization is not the "norm," variations in degree of nasalization and in the perceptual salience of the conditioning environment for vowel nasalization are normal consequences of speech production and perception and as such are the raw material of nasal vowel height shifts. Thus, we are brought, from another direction, to recognize the importance of variation in accounting for sound change (cf. Weinreich, Labow, & Herzog, 1968). It should be clear, however, that our acoustic-perceptual account of phonological changes in nasal vowel height has been restricted to the initiation of these changes. We have not attempted to specify the processes by which listener misperceptions become stable phonological patterns.

We have argued that listener familiarity with a particular phonetic and phonological structure leads to certain expectations with respect to vowel nasalization. Listeners correctly assess the contribution of nasal coupling to the vowel spectrum when these expectations are met, but when they are not, listeners apparently choose tongue configuration as an alternative source of the spectral effects of nasal coupling and thereby misperceive nasal vowel height. We conclude, then, that a comprehensive explanation of sound change in a language must take into account not only the physical (articulatory, acoustic, or perceptual) origins of the change, but also the phonetic and phonological structure of the language, including variability in that structure.

References

Abramson, A. S., Nye, P. W., Henderson, J., & Marshall, C. W. (1981). Vowel height and the perception of consonantal nasality. Journal of the Acoustical Society of America, 70, 329-339.

Ali, L., Gallagher, T., Goldstein, J., & Daniloff, R. (1971). Perception of coarticulated nasality. Journal of the Acoustical Society of America, 49, 538-540.

Beddor, P. S. (1983). Phonological and phonetic effects of nasalization on vowel height. Bloomington: Indiana University Lingustics Club.

Beddor, P. S. (1984). Formant integration and the perception of nasal vowel height. Haskins Laboratories Status Report on Speech Research, SR-77/78, 107-120.

Bedrov, Ya. A., Chistovich, L. A., & Sheikin, R. L. (1978). Frequency position of the 'center of gravity' of formants as a useful feature in vowel perception. Soviet Physics Acoustics, 24, 275-278.

Benguerel, A.-P., Hirose, H., Sawashima, M., & Ushijima, T. (1977). Velar coarticulation in French: a fiberscopic study. Journal of Phonetics, 5, 149-158.

Bhat, D. N. S. (1975). Two studies on nasalization. In C. A. Ferguson, L. M. Hyman, & J. J. Ohala (Eds.), Nasálfest: Papers from a symposium on nasals and nasalization (pp. 27-48). Stanford, CA: Stanford University.

Chen, M. (1971). Metarules and universal constraints on phonological theory. *Project on Linguistic Analysis* (University of California, Berkeley), *13*, MC1-MC56.

Chistovich, L. A., & Lublinskaya, V. V. (1979). The 'center of gravity' effect in vowel spectra and critical distance between the formants. *Hearing Research*, *1*, 185-195.

Chistovich, L. A., Sheikin, R., & Lublinskaya, V. (1979). 'Centres of gravity' and spectral peaks as the determinants of vowel quality. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 143-157). New York: Academic Press.

Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, *4*, 337-351.

Durand, M. (1956). Du rôle de l'auditeur dans la formation des sons du langage. *Journal de Psychologie Normale et Pathologique*, *52*, 347-355.

Entenman, G. (1977). *The development of nasal vowels*. Doctoral dissertation, University of Texas, Austin. (*Texas Linguistic Forum*, *7*.)

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Ferguson, C. A. (1963). Assumptions about nasals. In J. H. Greenberg (Ed.), *Universals of language* (pp. 53-60). Cambridge: MIT Press.

Foley, J. (1975). Nasalization as universal phonological process. In C. A. Ferguson, L. M. Hyman, & J. J. Ohala (Eds.), *Nasálfest: Papers from a symposium on nasals and nasalization* (pp. 197-212). Stanford, CA: Stanford University.

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, *24*, 127-139.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*, 386-412.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, *36*, 359-368.

Fujimura, O., & Lindqvist, J. (1971). Sweep-tone measurements of vocal-tract characteristics. *Journal of the Acoustical Society of America*, *49*, 541-558.

Haden, E. F., & Bell, E. A., Jr. (1964). Nasal vowel phonemes in French. *Lingua*, *13*, 62-69.

Haudricourt, A. G. (1947). *En/an* en français. *Word*, *3*, 39-47.

Hawkins, S., & Stevens, K. N. (1985). Acoustic and perceptual correlates of the non-nasal - nasal distinction for vowels. *Journal of the Acoustical Society of America*, *77*, 1560-1575.

Henderson, J. B. (1984). *Velopharyngeal function in oral and nasal vowels: A cross-language study*. Unpublished doctoral dissertation, University of Connecticut, Storrs.

House, A. S., & Stevens, K. N. (1956). Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders*, *21*, 218-232.

Hyman, L. M. (1972). Nasals and nasalization in Kwa. *Studies in African Linguistics*, *3*, 167-205.

Jonasson, J. (1971). Perceptual similarity and articulatory reinterpretation as a source of phonological innovation. *Speech Technology Laboratory Quarterly Progress and Status Report* (Royal Institute of Technology, Stockholm), *1*, 30-42.

Kawasaki, H. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 81-103). New York: Academic Press.

Krakow, R. A., Beddor, P. S., Goldstein, L. M., & Fowler, C. A. (in preparation). *Coarticulatory influences on perceived nasal vowel height.*

Kunisaki, O., & Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Research Institute for Logopedics and Phoniatrics Annual Bulletin* (University of Tokyo) *11*, 85-91.

Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal tract shapes from formant frequencies. *Journal of the Acoustical Society of America*, *64*, 1027-1035.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36.

Lightner, T. (1970). Why and how does vowel nasalization take place? *Papers in Linguistics*, *2*, 179-226.

Maddieson, I. (1984). *Patterns of sounds*. New York: Cambridge University Press.

Malécot, A. (1960). Vowel nasality as a distinctive feature in American English. *Language*, *36*, 222-229.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*, 213-228.

Martinet, A. (1955). *Economie des changements phonétiques*. Berne: A. Francke.

Martinet, A. (1965). Les voyelles nasales du français. *Linguistique*, *2*, 117-122.

Moll, K. L. (1962). Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research*, 5, 30-37.

Mrayati, M. (1975). Etude des voyelles françaises. *Bulletin de l'Institut de Phonétique de Grenoble*, *4*, 1-26.

Ohala, J. J. (1971a). Monitoring soft palate movements in speech. *Project on Linguistic Analysis* (University of California, Berkeley), *13*, J01-J015.

Ohala, J. J. (1971b). The role of physiological and acoustic models in explaining the direction of sound change. *Project on Linguistic Analysis* (University of California, Berkeley), *15*, 25-40.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (Eds.), *Historical linguistics II: Theory and description in phonology* (pp. 353-389). Amsterdam: North Holland Press.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1983, October). *Phonological evidence for top-down-processing in speech perception*. Paper presented at the Symposium on Invariance and Variability in Speech, MIT.

Ohala, J. J., Kawasaki, H., Riordan, C., & Caisse, M. (in preparation). *The influence of consonant environment upon the perception of vowel quality*.

Pandey, P. (1978). A physiological note on the effect of nasalization on vowel height. *International Journal of Dravidian Linguistics*, *7*, 217-222.

Passy, P. (1890). *Etude sur les changements phonétiques*. Paris: Librairie Firmin-Didot.

Paul, H. (1890/1970). *Principles of the history of language*. (Translated by H. A. Strong.) College Park, MD: McGrath Publishing.

Pope, M. K. (1934). *From Latin to Modern French*. Manchester: Manchester University Press.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, *70*, 321-328.

Ruhlen, M. (1978). Nasal vowels. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language*: *Vol. 2.*: *Phonology* (pp. 203-242). Stanford: Stanford University Press.

Schourup, L. (1973). A cross-language study of vowel nasalization. *Ohio State Working Papers in Linguistics*, *15*, 190-221.

Stevens, K. N., Fant, G., & Hawkins, S. (forthcoming). Some acoustical and perceptual correlates of nasal vowels. In R. Channon & L. Shockey (Eds.), *Festschrift for Ilse Lehiste*. Dordrecht, Holland: Foris.

Straka, G. (1955). Remarques sur les voyelles nasales, leur origine et leur evolution en français. *Revue de Linguistique Romane*, *19*, 245-274.

Sweet, H. (1888). *History of English sounds*. Oxford: Oxford University Press.

Ushijima, T., & Sawashima, M. (1972). Fiberscopic observation of velar movements in speech. *Research Institute of Logopedics and Phoniatrics Annual Bulletin* (University of Tokyo), *6*, 25-38.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 97-195). Austin: University of Texas Press.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America*, *69*, 275-282.

Wright, J. (1980). The behavior of nasalized vowels in the perceptual vowel space. *Report of the Phonology Laboratory* (University of California, Berkeley), *5*, 127-163.

Footnotes

[1] The F2 differences in [e] and [ẽ] in Figure 4 suggest that the two vowels may have been produced with different oral tract configurations, thus we cannot say to what extent the shift in center of gravity is due to nasalization per se.

[2] Although English does not typically have distinctive nasal vowels before voiced stops, an apparent exception to nondistinctive vowel nasalization in American English occurs before voiceless stops. Malécot (1960) found that nasal consonants before voiceless stops are of extremely short duration, and may possibly be absent for some speakers, suggesting the existence of minimal pairs (e.g., *cat* versus *can't*) differing only in vowel nasality.

[3] For some discussion of factors leading to weakening of nasal consonants, see Lightner (1970), Ohala (1971b), Schourup (1973), Foley (1975), Entenman (1977), and Ruhlen (1978).

THE THAI TONAL SPACE*

Arthur S. Abramson†

In the analysis of a tone language, the linguist normally thinks first of pitch levels and glides as the probable phonetic basis of phonologically relevant tones. This is true even though there may be other features, apparently secondary in importance, that go along with pitch. Of course, it is well known that in some languages, as in certain dialects of Vietnamese, a feature other than pitch may be dominant in one or more of the tones.

Against the background of earlier auditory (e.g., Haas & Subhanka, 1945) and instrumental (Bradley, 1911) analysis, Abramson (1962) was apparently the first to combine techniques of acoustic analysis and speech synthesis to investigate the tones of Central Thai (Siamese)--or, indeed, any tone language--both acoustically and perceptually. Since then, of course, other such treatments of Asian languages, including Thai, have appeared (e.g., Gandour, 1978).

The present study is part of an ongoing exploration (e.g., Abramson, 1975, 1976) of the Thai tonal "space." This space is taken to be the set of articulatory and auditory dimensions by which the speaker is constrained in production and perception. The paper makes use of unpublished or reanalyzed data obtained in Thailand from time to time at the old Central Institute of English Language at Mahidol University, the Faculty of Humanities of Ramkhamhaeng University, and the Faculty of Arts of Chulalongkorn University. It has three broad goals: to revalidate earlier work on "ideal" contours for the tones on isolated monosyllables, to gain some insight into the latitudes of shifting levels and glides for the intelligibility of the tones, and to take another look (cf. Abramson, 1978) at the typological usefulness of the distinction between static and dynamic tones.

The identifiability of isolated natural Thai tones had been demonstrated in Abramson (1962) and was reaffirmed with much more extensive testing in Abramson, 1975. These findings were a necessary precursor to the five experiments with synthetic tones presented in this report. Aside from the baseline data for all five tones obtained in Experiment 1, the report gives no

---

*To appear in the Proceedings of the 18th International Conference on Sino-Tibetan Languages and Linguistics, Bangkok, August 27-29, 1985.

†Also University of Connecticut

Acknowledgment. This work was supported NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. The American Council of Learned Societies and the University of Connecticut Research Foundation enabled the author to present an oral version in Bangkok at the 18th International Conference on Sino-Tibetan Languages and Linguistics.

serious attention to the falling tone, which will have to be treated in another paper.

Experiment 1

The major physical correlate of the psychological feature pitch is **fundamental frequency** (F_0), which, for speech, varies with the vibration rate of the larynx. The speech synthesizer used in Abramson (1962) has long since gone out of use. For this experiment, and the rest, the Haskins Laboratories computer-controlled formant synthesizer was used. The syllable specified segmentally as [k^ha:] was chosen as the carrier for the five tones of Central Thai, yielding five tonally differentiated words. Each synthetic syllable was made 450 ms long. The frequencies and amplitudes of three steady-state formants, simulating resonances of an adult male vocal tract, were made appropriate for a vowel of the type [a:], with formant transitions that yielded the percept of an initial dorso-velar stop. Timing of the source functions was set to produce a voiceless aspirated stop. This was done by turning on a turbulent source for the first 80 ms of the pattern (Lisker & Abramson, 1970), followed by a periodic buzz source to simulate glottal pulsing for the remaining 370 ms; the latter served as the carrier for the F_0 contours. A slight upward tapering of the overall amplitude at the beginning and a slight downward one at the end made for greater naturalness.

For Experiment 1, the five F_0 contours (Figure 1) found in Abramson (1962) to be ideal for the synthesis of the tones were replicated as closely as possible with the newer synthesizer and imposed on tokens of the carrier syllable. These were played in a number of random orders, over the period of a month, to 37 native speakers of Central Thai, who wrote their responses as words in Thai script. The results, given in Figure 2, reveal rather robust identification functions. The two least satisfactory percepts are the mid and low tones, although both contours do achieve 88% identification. The falling, high, and rising tones are at least 10% higher. All three of them, including the allegedly static high tone, involve much F_0 movement.

Experiment 2

In this experiment and in the remaining three, simple straight-line contours were used for a partial exploration of the tonal space. The 16 contours prepared for Experiment 2 are shown in Figure 3. These variants all start at 106 Hz, the top of the lower third of the voice range, and go to endpoints ranging from 90 to 152 Hz in 4-Hz steps. (An accidental exception is a 6-Hz step from 106 to 112 Hz.)

Four hypotheses were put forth: (1) The beginning portion of this fanlike array is too low in the voice range for mid-tone responses. (2) The falls at the lower part of the array are too low and slow for the falling tone. (3) The upper variants rise too slowly for the rising tone. (4) The labels used for the set by the subjects should be mainly "low" and "high."

The responses to the stimuli are given in Figure 4. The first hypothesis is weakly confirmed in that the mid tone has a peak, just for the level variant at 106 Hz, of only 39%. The second hypothesis is confirmed; the word with the falling tone is not used as a label at all. The third hypothesis is not well supported, since the highest variant is labeled "rising" 64% of the time; however, this peak, with only two variants above 50%, is not very robust

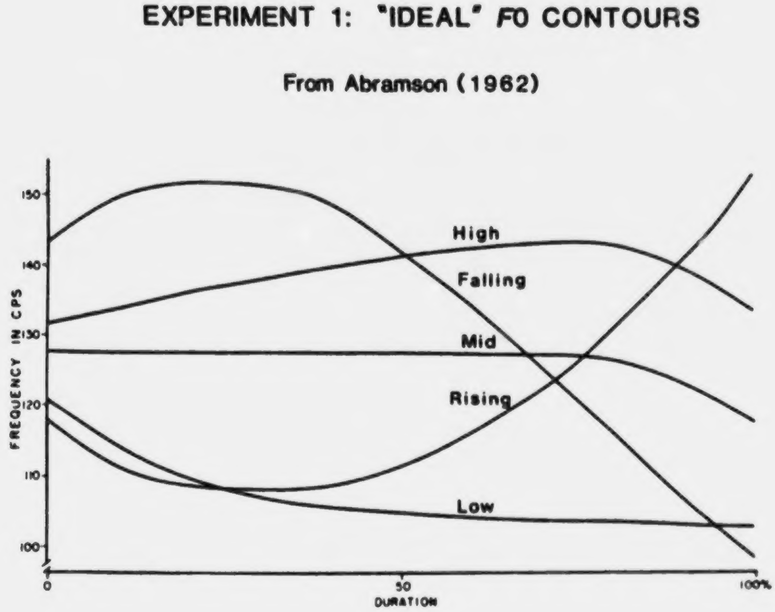


Figure 1. F_0 contours for the Thai tones of an adult male on long vowels resynthesized from Abramson (1962: Figure 3.6).

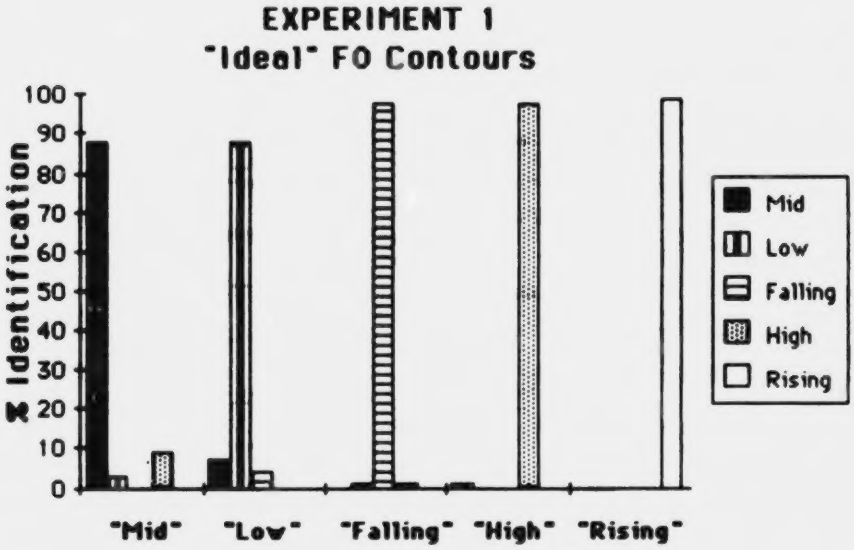


Figure 2. Experiment 1: Identification of the contours of Figure 1 by 37 subjects.

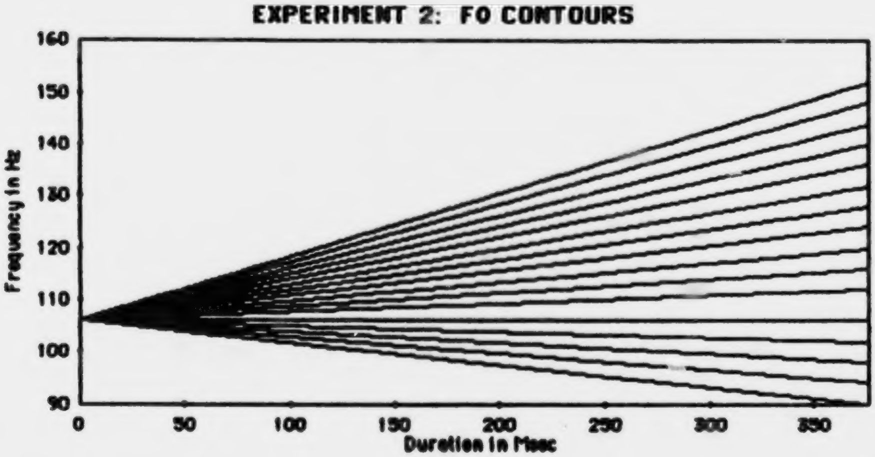


Figure 3. Sixteen F_0 contours moving from 106 Hz to endpoints ranging from 90 to 152 Hz.

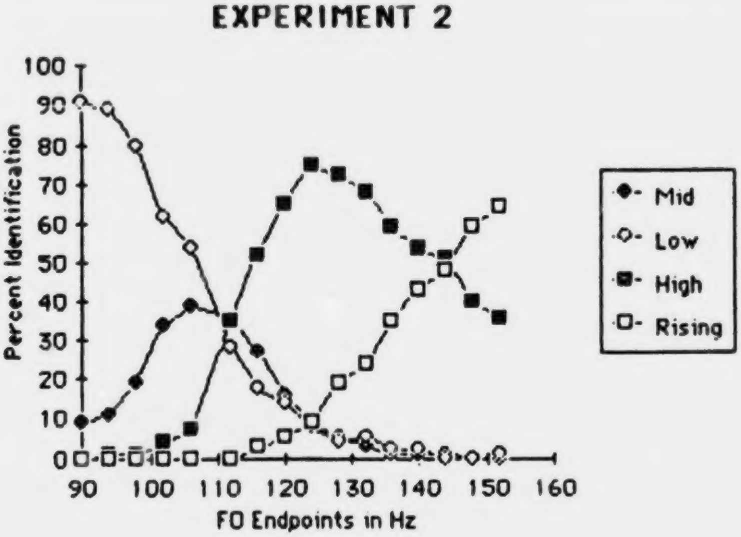


Figure 4. Experiment 2: Identification of the contours of Figure 3 by 38 subjects.

compared with the high tone, which has seven variants above 50%, and the low tone, which has five variants above 50% and a peak at 90%. As for the fourth hypothesis, it is true that the major peaks in the figure are for the low and high categories, but the labeling function for the rising tone is conspicuous too, while the mid tone, reaching a peak of 39%, is at least not negligible.

Experiment 3

Figure 5 shows the stimuli for this experiment. They are 17 F_0 contours on tokens of the [kʰa:] carrier syllable. The contours all start at 90 Hz, the bottom of the simulated voice range, and go to endpoints ranging, once again, from 90 to 152 Hz in 4-Hz steps. (The exception is the first step, which is from 90 to 92 Hz.) The original intent had been to make 92 Hz the bottom frequency.

This array was meant to explore four hypotheses: (1) The onsets are too low in the voice range to yield the mid tone. (2) The low onsets should give a much better rising category than in Experiment 2. (3) There should be no high-tone responses. (4) The first two or three contours at the bottom ought to be heard mainly as the low tone.

The results of Experiment 3 are given in Figure 6. With the labeling function of the mid tone hovering around 10% over the first half of the stimulus array and then dropping to nothing, the first hypothesis is well supported.

The rising-tone category is clearly more robust here than in Experiment 2, thus confirming the second hypothesis. More abrupt rises to the same endpoints produce more convincing tokens of the rising tone. Although the labeling function for the high tone is rather poor, with a plateau at about 40% for four of the stimuli, this result does not bear out the very categorical prediction of the third hypothesis. Of course, this should be compared with Experiment 2 in which the higher starting point led to a much more robust high-tone percept. In agreement with the fourth hypothesis, the first few contours are heard predominantly as the low tone; however, the greater area under the "low" curve in Experiment 2 (see Figure 4) suggests that a slight fall enhances the acceptability of those stimuli.

Experiment 4

This time, the full voice range furnishes the set of beginning points and the top of the range, the endpoint. Thus, as shown in Figure 7, the beginnings of the 16 contours range from 90 to 152 Hz in 4-Hz steps, except for a 5-Hz step at the bottom (90 to 95) and a 3-Hz step at the top (149 to 152). All the contours end at 152 Hz.

The hypothesis here is that only the high and rising tones should be heard. This portion of the tonal space seems utterly unsuitable for any other tone.

In fact, aside from the essentially negligible "low" labels along the bottom of the graph in Figure 8, the two categories that emerge are the high and rising tones. Interestingly enough, the stronger of the two categories is the high tone. Apparently, these less abrupt rises, compared with those of Experiment 3 (Figures 5 and 6), bias the response toward the high tone.

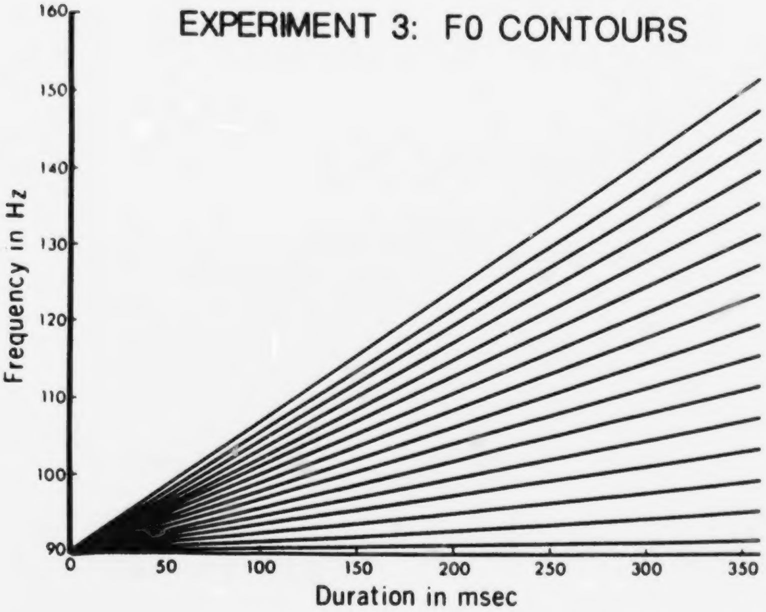


Figure 5. Seventeen F_0 contours moving from 90 Hz to endpoints ranging from 90 to 152 Hz.

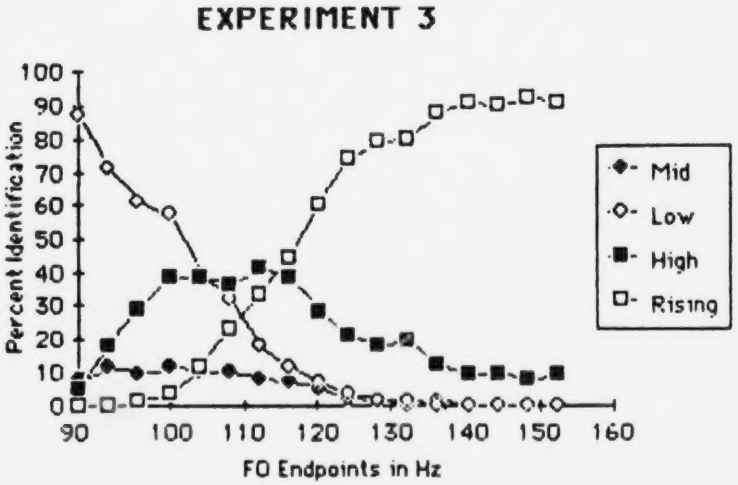


Figure 6. Experiment 3: Identification of the contours of Figure 5 by 38 subjects.

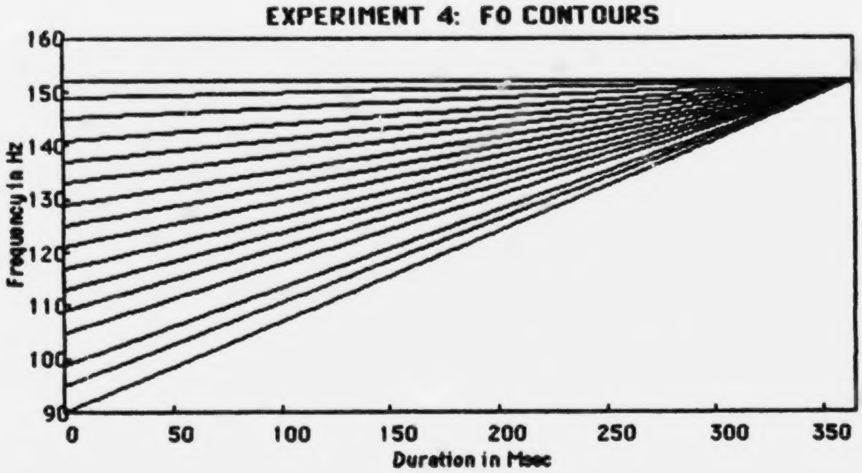


Figure 7. Sixteen F_0 contours starting at points ranging from 90 to 152 Hz, all ending at 152 Hz.

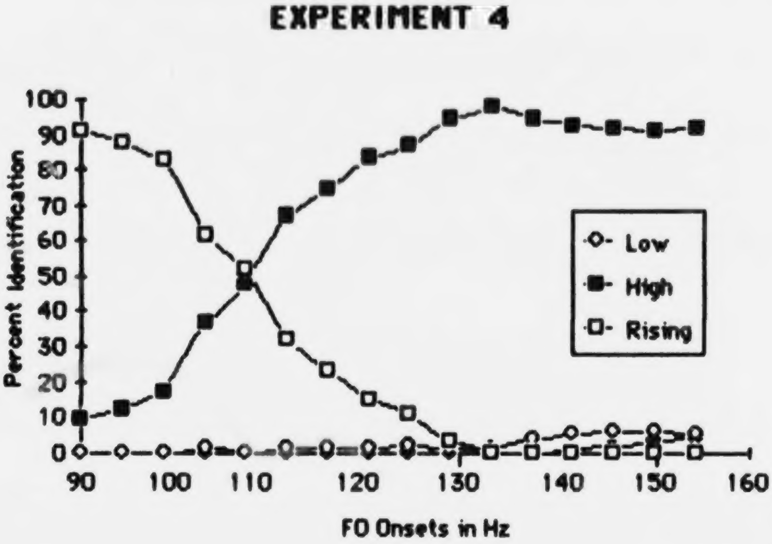


Figure 8. Experiment 4: Identification of the contours of Figure 7 by 38 subjects.

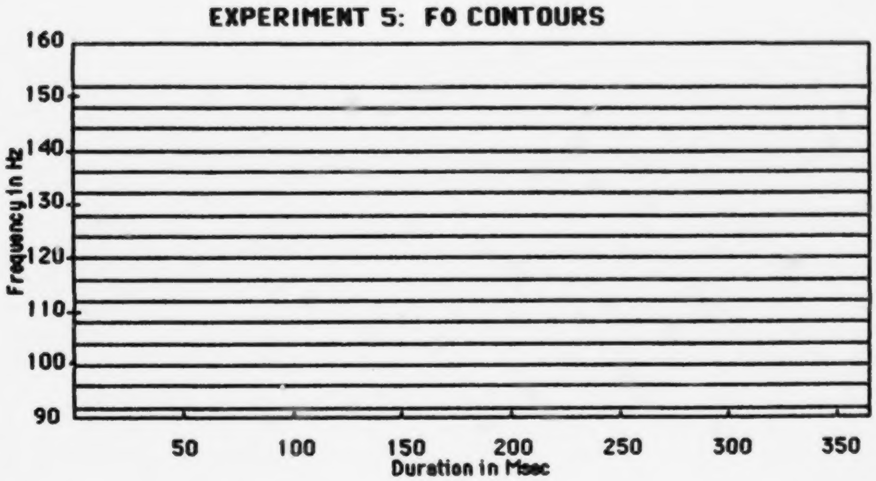


Figure 9. Sixteen level F_0 contours ranging from 92 to 152 Hz.

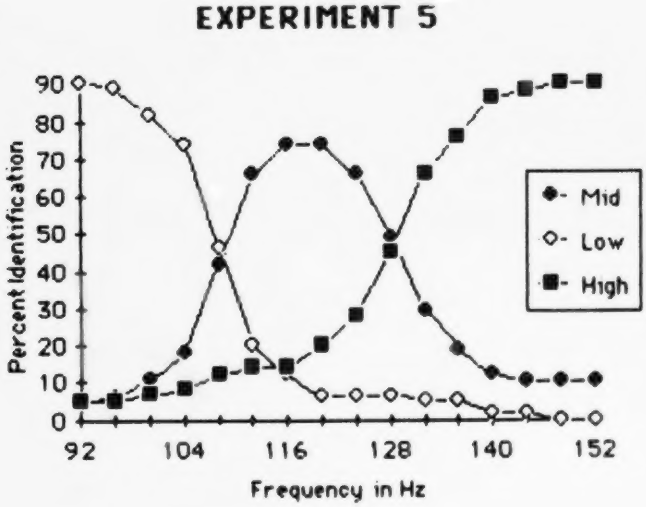


Figure 10. Experiment 5: Identification of the contours of Figure 9 by 37 subjects (adapted from Figure 2 in Abramson 1978).

Experiment 5

Here, on tokens of synthetic [k^ha:], there are 16 level contours, ranging from 92 to 152 Hz in 4-Hz steps, as seen in Figure 9. These are undoubtedly a greater deviation from natural speech than any of the foregoing contours; nevertheless, given the frequent assumption of "level" tones in the linguistic literature, it was important to see what the perceptual response to such stimuli would be. Indeed, the hypothesis expected only static tones, that is, the mid, low, and high tones.

The results, first presented in Abramson (1978), are given in Figure 10. Only the mid, low, and high categories appear. There is much overlap, resulting in a lower peak for the mid tone than for the other two.

Conclusion

This study continues to support the primacy of the fundamental frequency of the voice as the carrier of tonal information in Thai, although some concomitant features may, in certain contexts, have at least secondary cue value. The "ideal" contours found in earlier work (Abramson, 1962; Erickson, 1974; Gandour, 1975) are still quite acceptable for isolated Thai words.

The new work has yielded some information on the perceptual latitudes of four of the tones. Level contours are fairly good for the static tones. For absolute levels to be so identified in citation forms of words in natural speech, there must be some auditory accommodation to the speaker's voice range (Abramson, 1976; Leather, 1983), as well as to the immediate tonal context. A comparison of Figures 8 and 10 does reveal, however, that the high-tone percept is improved by F_0 movement. (Similar observations were made for the mid and low tones in Abramson, 1978.)

Fairly rapid movements are needed for the dynamic tones. This conclusion is supported here only for the rising tone, although data not presented here show the same effect for the falling tone. While the dichotomy between static and dynamic tones is thus not categorical, it does have some perceptual support.

There is more work to be done on the tonal space for Thai and other languages. The present findings seem compatible with the pitch features isolated by Gandour (1978) and the emphasis on the importance of the onset frequency values of the contours for Thai by Saravari and Imai (1983). Of course, in running speech, all this is further complicated by interactions between sentence intonation and the tonal space (Abramson & Svastikula, 1983).

References

Abramson, A. S. (1962). The vowels and tones of standard Thai: Acoustical measurements and experiments. Bloomington, IN: Indiana Univ. Research Center in Anthropology, Folklore, and Linguistics, Pub. 20.

Abramson, A. S. (1975). The tones of Central Thai: Some perceptual experiments. In J. G. Harris & J. R. Chamberlain (Eds). Studies in Tai linguistics in honor of William J. Gedney (pp. 1-16). Bangkok: Central Institute of English Language.

Abramson, A. S. (1976). Thai tones as a reference system. In T. W. Gething, J. G. Harris, & P. Kullavanijaya (Eds.), *Tai linguistics in honor of Fang-Kuei Li* (pp.1-12). Bangkok: Chulalongkorn University Press.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*, 319-325.

Abramson, A. S., & Svastikula, K. (1983). Intersections of tone and intonation in Thai. *Haskins Laboratories Status Report on Speech Research*, *SR-74/75*, 143-67.

Bradley, C. B. (1911). Graphic analysis of the tone-accents of the Siamese language. *Journal of the American Oriental Society*, *32*, 282-289.

Erickson, D. (1974). Fundamental frequency contours of the tones of standard Thai. *Pasaa*, *4*, 1-25.

Gandour, J. (1975). On the representation of tone in Siamese. In J. G. Harris & J. R. Chamberlain (Eds), *Studies in Tai linguistics in honor of William J. Gedney* (pp. 170-195). Bangkok: Central Institute of English Language.

Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41-76). New York: Academic Press.

Haas, M. R., & Subhanka, H. R. (1945). *Spoken Thai*. New York: Henry Holt.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics*, *11*, 373-382.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress Phonetic Sciences Prague 1967* (pp. 563-567).

Saravari, C., & Imai, S. (1983). Perception of tone and short-long judgement of vowel of variants of a Thai monosyllabic sound. *Journal of Phonetics*, *11*, 231-242.

P-CENTERS ARE UNAFFECTED BY PHONETIC CATEGORIZATION*

André Maurice Cooper,† D. H. Whalen, and Carol Ann Fowler††

Abstract. The perceived onset (P-center) of a word typically does not correspond to its acoutsic onset (Marcus, 1981; Morton, Marcus & Frankish, 1976). Some researchers have suggested that the P-center of a word is solely a product of the acoustic characteristics of the word, while other have suggested that a word's P-center is determined by its phonetic characteristics. The present series of experiments pits a continuously varying acoustic parameter against a categorical phonetic percept in order to determine whether P-center location is sensitive to the phonetic identity of the prevocalic segments of a syllable. With a /ša/-/ča/-/ta/ continuum and three different /sa/-/sta/ continua, we find that phonetic judgments are categorical but P-center judgments are continuous. The results demonstrate that P-center location is not determined by the phonetic identity of syllable initial consonants. Nor, however, is it determined by the rise time or the amplitude envelope of the signal as Howell (1984) has suggested. Instead, as Morton et al. and Marcus recognized, a combination of at least two different parts of the signal is at work, namely, the duration of the prevocalic consonant or consonants and, to a lesser extent, the duration of the syllable rhyme. Whereas the relevant dimension of each component of the syllable is duration, acoustically defined, the partitioning of the syllable is phonetically motivated. Thus, both the phonetic structure of a syllable and the particular acoutic realizations of its structure affect the location of the P-center.

When listeners are presented with sequences of consonant-vowel syllables differing in the number or the nature of consonants and with equal intervals between their acoustic onsets, they judge the rhythm of the sequences to be irregular. Furthermore, when given the opportunity to adjust the relative timing of two syllables until they are perceptually isochronous, listeners introduce systematic deviations from acoustic onset isochrony (Morton et al., 1976). These deviations can not be explained by reference to any obvious acoustic events such as the peak intensity of an utterance or the acoustically marked onset of the stressed vowel (see Allen, 1972; Morton et al., 1976; Rapp, 1971).

These findings indicate that the event that listeners attend to when judging relative timing is opaque to conventional measurement techniques.

---

*Perception & Psychophysics, 1986, 39, 187-196.

†Also Department of Linguistics, Yale University.

††Also Department of Psychology, Dartmouth University.

Acknowledgment. This research was supported by NIH Grant 16591, NIH Grant HD 01994, and NSF Grant BNS 8111470 to Haskins Laboratories.

Therefore, Morton et al. (1976, p. 405) do not attempt to locate the event absolutely, but propose that listeners base their rhythmicity judgments on a word's "psychological moment of occurrence" or its "P-center" (perceptual center). According to Marcus (1981), although the P-center of a word cannot be determined absolutely, it can be located relative to the timing of other speech or nonspeech events. Thus, by hypothesis, in order for two syllables presented in continuous alternation to be perceived as isochronous, the components of the sequence must have their P-centers at equal intervals.

Syllables whose initial consonants differ in manner generally have different P-center locations (Fowler & Tassinary, 1981). These consonants, in turn, have different acoustic characteristics, especially in the duration of the signal before the first vocalic pitch period (loosely, "consonant duration"). Marcus (1981) found that the location of the P-center is highly correlated with initial consonant duration. Specifically, the shorter the duration of the consonant, the earlier the P-center with respect to the acoustic onset of the syllable (also see Rapp, 1971; Fowler & Tassinary, 1981). Marcus also found that changes in the duration of segments following the initial consonants are associated with a smaller yet significant change in the location of the P-center. These two relationships are expressed by the equation:

$$P = .65x + .25y + k,$$

where x is the measured duration of a syllable onset (that is, the part of the signal preceding the first oral pitch pulse), y is the duration of the syllable rhyme (the vocalic segment and final consonants) and k is an arbitrary constant reflecting the fact that the equation predicts the relative, rather than the absolute, location of the P-center. Although the equation accounts for about 90% of the variability of P-center locations in the set of digits one to nine, it does not explain this variability.

In fact, one issue that the equation leaves in question is whether P-center shifts are explained by the phonetic or by the acoustic properties of a word. According to Marcus's equation, P-centers have a phonetic basis to the extent that the equation predicts that syllable-initial consonants have a markedly greater affect on P-center location than do vowels (whether syllable initial or nonsyllable initial) or final consonants. However, Marcus also shows that durational changes that do not affect the phonetic identity of the segments in a word do affect P-center location. Accordingly, the P-center is not solely a product of the phonetic identity of a segment.

Marcus attempted to test the effect of phonetic identity on P-center location by pairing the members of a phonetic continuum with several reference stimuli and adjusting the relative timing of the stimulus pairs to isochrony. The continuum was created by deleting successive portions of the /s/-noise (in 30-ms decrements) from a naturally-produced token of the word /sɛvən/. The initial consonant of the continuum stimuli spanned three phonetic categories, viz., /s/, /ts/ and /d/ (presumably as judged by Marcus, 1976, himself). Marcus found that abrupt shifts in phone categorization across the phonetic continuum were accompanied by a continuous change in P-center location across the continuum. Marcus's continuum, however, was not sufficiently well constructed to address the issue adequately. First, no identification or discrimination tests were performed to confirm how the continuum was perceived. Second, the steps of Marcus's continuum were so large that the initial consonants spanned three phonetic categories (including one, /ts/, that is not phonotactically possible in English) within four steps of the five-step continuum.

In the present study, we used a series of phonetic continua to investigate the extent to which the phonetic realization of a syllable-initial consonant is relevant to P-center location. Our experiments improved upon and extended Marcus's experiment in several ways. First, we obtained identification and discrimination data to confirm that the continuum stimuli were categorically perceived. Second, the acoustic differences between neighboring stimuli on the continuum were sufficiently small to enable us to address the question of the relationship between phonetic consonantal categories and the location of the P-center. Third, all of the phonetic categories under investigation, /š/-/č/-/t/ and /s/-/st/, occur syllable-initially in English, the native language of the listeners. The phonetic categories /s/ and /st/ had already been shown to have P-center differences on the order of 26 ms (Fowler & Tassinary, 1981). Finally, we manipulated both prevocalic and postvocalic durations, allowing us to test Marcus's equation directly. Our results also allowed us to address Howell's (1984) recent suggestion that the amplitude envelope of a syllable significantly affects the location of its P-center.

We compared the phonetic categorization of syllable initial consonants and the relative location of P-centers to determine their effects upon each other. The test stimuli consisted of a /ša/-/ča/-/ta/ continuum and three /sa/-/sta/ continua. The construction of our first continuum was guided by the fact that when a sufficient amount of frication is deleted from /ša/, listeners hear /ča/ rather than /ša/; as additional frication is deleted, listeners eventually hear /ta/. The construction of the remaining continua was based on the fact that when a sufficiently long silent gap is introduced between the frication and the vocalic segment of the /sa/ syllable, listeners hear /sta/ rather than /sa/.

Repp (1984) identified four criteria that delimit categorical perception: there must be (1) an abrupt shift in labeling probabilities somewhere along the continuum, (2) a peak in the discrimination function at the category boundary, (3) chance or near-chance level discrimination of stimuli within categories, and (4) perfect predictability of the discrimination function from the identification function. Strict categorical perception, as described above, is rarely, if ever, reported in the literature; instead, the actual data approximate the ideal more or less well. Repp (1984) emphasizes that provided that the other criteria are not severely violated, a peak in the discrimination function at the category boundary is the crucial defining characteristic of categorical perception.

Three relationships between the categorical nature of the continua and the P-center function are possible. First, the relative location of the P-center could be sensitive only to the phonetic properties of the syllables. In this case, listeners' P-center judgments would be predictable from their identification functions. For syllables with categorically-perceived consonants, this would imply negligible within-category P-center shifts, but noticeable shifts between categories. Second, the P-center could be sensitive to the durations of a syllable's acoustic segments. In this case, listeners' P-center judgments would vary monotonically as a function of duration (see Marcus, 1981). The third possibility is that P-center judgments would be influenced by both phonetic and acoustic properties of the stimuli, resulting in both an abrupt shift in the P-center at the category boundary and systematic variation within categories.

General Procedures, Experiments 1-4

Each experiment consisted of three tasks. The first two tasks, a forced-choice identification test and an AXB discrimination test, were used to determine the extent to which each continuum was categorically perceived. For the identification test, multiple repetitions of the stimuli were randomized and presented to listeners. For the AXB test, multiple repetitions of all pairs of syllables in a continuum differing by one step (Experiment 1) or by two steps (Experiments 2 - 4) were randomized and presented to listeners.

The final task, the alignment test, was designed to measure the relative P-center location of the test stimuli. For the alignment test, each of the test stimuli was paired with a reference syllable /ba/ for presentation. (The reference syllable was 329 ms in duration). The syllable pairs were played to listeners in a continuous sequence under computer control. The temporal position of the second syllable relative to the first was adjustable within a window of fixed duration. Initially, on each trial, there was a 50-ms gap between offset of the fixed syllable and the onset of the movable stimulus. The listener's task was to adjust the timing of the sequence until it was perceived as isochronous. When the listener was satisfied with the adjustment, the computer reported the interval between the acoustic onsets of the stimuli. Two systems, with minor differences, were used. On one system, implemented with a New England Digital computer at Dartmouth College, listeners adjusted the second syllable in steps of 15 ms, 5 ms or 1 ms in either direction relative to the fixed syllable by pressing designated keys on a computer terminal keyboard. On the other system, implemented with a DEC GT40 computer at Haskins Laboratories, the alignments were made by turning a knob. The analog output of the knob was digitized to indicate adjustments in 12.8-ms increments. Since the experimental results obtained from the two systems were similar, they were combined.

Experiment 1

The purpose of the first experiment was to investigate how the location of P-centers might vary across a categorically perceived /ša/-/ča/-/ta/ continuum.

Method

Stimuli. Using a waveform editor, a 10-step /ša/-/ča/-/ta/ continuum was created by deleting 15-ms increments of frication from the acoustic onset of a digitized naturally spoken /ša/ syllable. The fricative segment of the original /ša/ was 189 ms in duration; the vocalic segment was 266 ms in duration. Syllable duration covaried with the duration of /š/ in the continuum. To minimize abrupt onsets, an amplitude ramp, linear with sound pressure, was applied to the onset of each stimulus. The offset of the ramp was fixed at 150 ms into the frication of the original /ša/, while the total duration (and steepness) of the ramp varied with the duration of the frication. Thus, for one extreme of the continuum, the ramp was applied to the initial 150 ms of the original stimulus. For the other extreme, the first 135 ms of noise was deleted from the frication and a linear taper was applied to the initial 15 ms of the frication (see Figure 1). Although the onset ramps in our continua do become steeper as the continuum stimuli get shorter, this seemed to us an improvement over the procedure of Marcus (1981), who made no attempt to avoid abrupt onsets. In his continuum, increasing amounts of fricative energy were simply removed from the beginning of the word "seven," resulting in abrupt onsets, and hence the perception of /ts/ very early in the continuum.

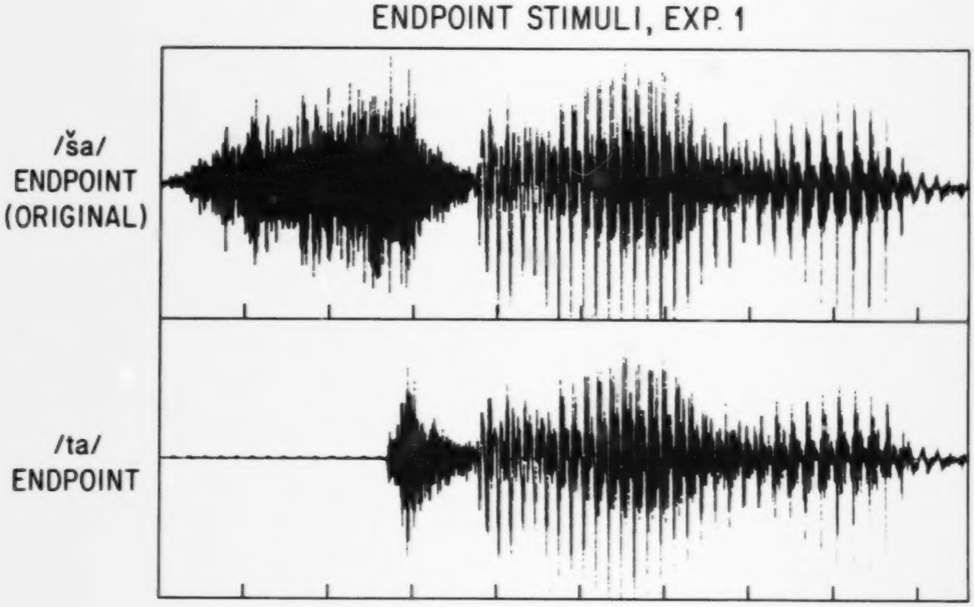


Figure 1. Continuum endpoint stimuli for Experiment 1: unmodified stimulus (upper panel), extreme stimulus manipulation (lower panel).

The identification test consisted of a randomized sequence of 10 repetitions of each member of the continuum (10 x 10 = 100 trials). The categories allowed in the identification test were /t/, /č/ and /š/. A one-step AXB discrimination test consisted of a randomized sequence of five repetitions of the four versions of each of the nine pairings of the stimuli (9 x 4 x 5 = 180 trials). For the alignment test, twenty judgments were obtained for each member of the continuum (20 x 10 = 200 trials).

Subjects. Three subjects participated. Two were naive as to the purposes of the experiment and the third was one of the authors (CAF).

Results and Discussion

Figure 2 shows the mean identification and discrimination functions for the three subjects. The ordinate represents the percent identification and the percent of correct discrimination. The abscissa represents the members of the continuum.

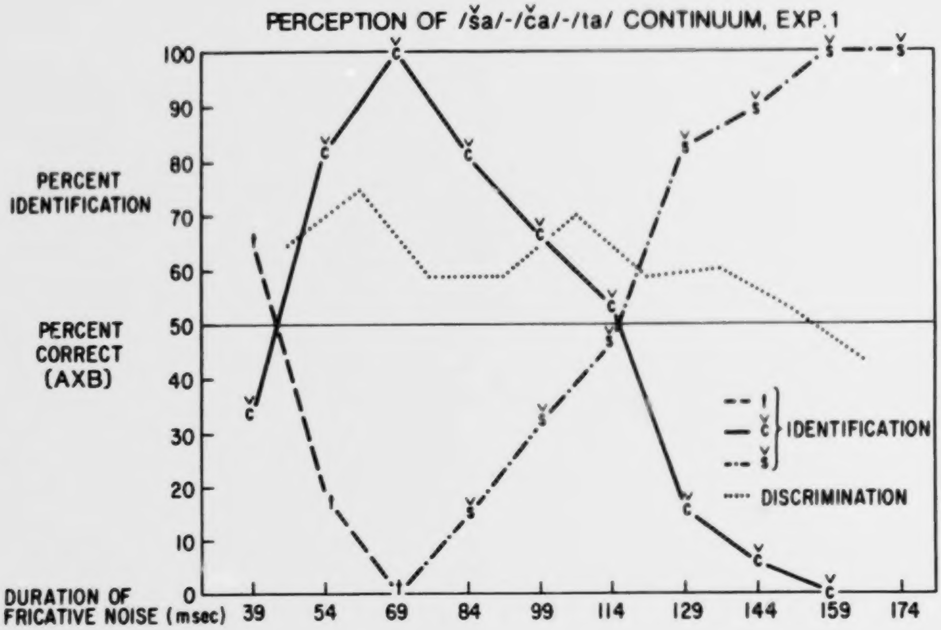


Figure 2. Mean identification and discrimination functions (pooled across three subjects) for /Ša/-/Ča/-/ta/ continuum, for Experiment 1.

The identification function shows two abrupt shifts in phoneme identification. The mean category boundaries (the 50% crossover points of the identification function) occur at 115 ms of frication for /Ša/-/Ča/ for all subjects and at 44 ms of frication for /Ča/-/ta/ for the two subjects who reported /ta/'s. The discrimination function shows peaks in discrimination near the mean category boundaries, indicating that listeners discriminate better between stimuli that straddle a phoneme boundary than between stimuli that fall within the same phoneme category. Our data show two departures from strict categorical perception that are frequently reported in the literature. First, our discrimination peak is slightly offset from the category boundary, and second, within-category discrimination is above chance level (see Best, Morrongiello, & Robson, 1981; Healy & Repp, 1982; Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Harris, Kinney, & Lane, 1961). Nevertheless, the pattern of our data is similar to the patterns obtained in earlier studies in which researchers concluded that perception was categorical.

The mean results of the three subjects' performance on the alignment test are shown in Figure 3. The ordinate represents the displacement from acoustic isochrony of the test stimuli relative to the reference syllable in ms (i.e., the interval from the acoustic onset of the test stimuli to the acoustic onset of the reference syllable minus one half the window size; thus, this measure would be zero for stimuli aligned at their acoustic onsets). The abscissa represents the duration of the fricative noise in ms.

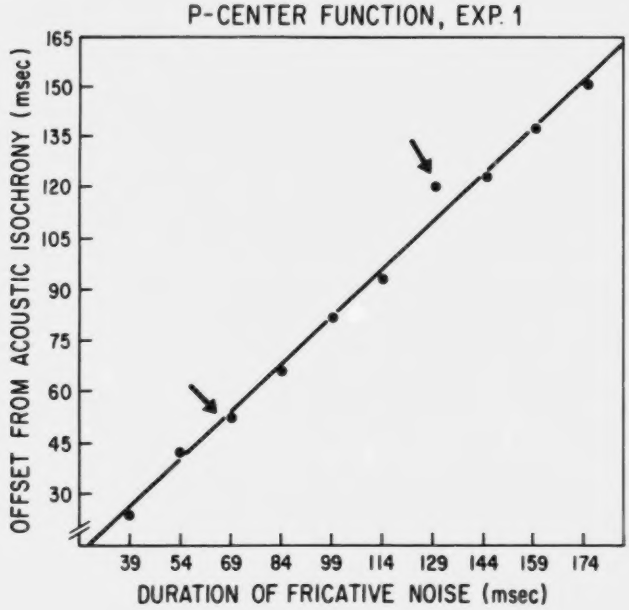


Figure 3. Mean P-center alignment function for Experiment 1, pooled across three subjects. The solid line represents the regression line and the arrows indicate the category boundaries.

In this experiment, P-center location, as shown by the regression line, moves linearly toward stimulus offset as the duration of the noise increases across the continuum. The slope of the regression line is .95, with $\underline{r}$ = .73. (The slopes of the P-center regression lines for the individual subjects are .95, 1.07, and .84; the correlations are .80, .91, and .57, respectively. Each correlation is significant at the .001 level). Thus, there is essentially a millisecond shift in P-center location for every millisecond of frication deleted.

There is no abrupt shift in P-center location at the category boundaries (which are indicated by the arrows). This indicates that the phonetic identity of syllable-initial consonantal segments does not affect P-center location. That there is no phonetic (or any other) source of nonlinearity in the data is revealed by a goodness-of-fit test. This test reveals that the first-degree polynomial is the highest one to significantly reduce the residual sum of squares. The significant $\underline{F}$, $\underline{F}(3,596) = 233.8$, $\underline{p} < .001$, for degree 0 indicates that there is systematic variance unaccounted for at that level; the nonsignificant $\underline{F}$ for degree 1, $\underline{F}(2,596) = .33$, n.s., indicates that the linear function accounts for as much of the variance as any higher polynomial. Thus, the point on the graph just after the category boundary does not signal a departure from linearity.

Experiment 2

The purpose of the second experiment was to replicate and extend the results of the first experiment using a /sa/-/sta/ continuum. In contrast to Experiment 1, the onset characteristics (including the rise time) of the stimuli in this experiment were left unaffected by the experimental manipulation. By eliminating the confounding effects of rise time, we are able to address Howell's (1984) claim that the amplitude envelope of a syllable (manipulated in his experiment by varying the syllable's rise time) significantly affects its P-center location.

Method

Stimuli. An eleven-step /sa/-/sta/ continuum was created by inserting 10-ms increments of silence between the frication and the vocalic segment of a naturally spoken /sa/ syllable. The frication of the original syllable was 206 ms in duration and the vocalic segment was 360 ms in duration. The first stimulus of the continuum was the original /sa/. The final stimulus contained a 100-ms gap between the frication and the vocalic segment (see Figure 4).

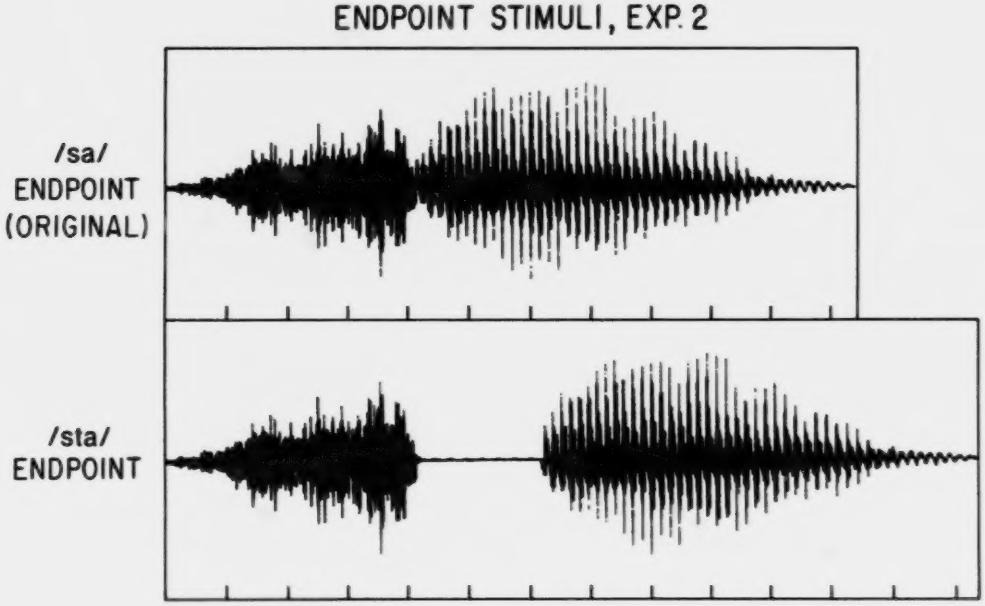


Figure 4. Continuum endpoint stimuli for Experiment 2: unmodified stimulus (upper panel), extreme stimulus manipulation (lower panel).

The identification test consisted of a randomized sequence of 20 repetitions of each member of the continuum (20 x 11 = 220 trials). A two-step AXB discrimination test consisted of a randomized sequence of four repetitions of the four versions of each of the nine pairings of the stimuli, within the AXB paradigm (9 x 4 x 4 = 144 trials). A two-step discrimination test was used rather than a one-step discrimination test (as in Experiment 1) because we found 10-ms differences between stimuli to be too small for listeners to to discriminate consistently. (The stimuli in the one-step comparisons in Experiment 1 differed by 15 ms.) For the alignment test, 12 judgments were obtained for each member of the continuum (12 x 11 = 132 trials).

Subjects. Three subjects participated in the experiment. One subject was naive as to the purposes of the experiment. The other two subjects were two of the authors (AMC and CAF).

Results and Discussion

Figure 5 shows the mean identification and discrimination results for the three subjects. The ordinate represents the percentage of /sta/ responses for the identification data and percent correct for the discrimination data. The abscissa represents gap duration. The identification function shows an abrupt shift in phoneme identification. The mean category boundary occurs at about 60 ms of silence. The discrimination function shows a peak in discrimination near the mean category boundary and troughs in performance within categories. (The high discriminability of the first stimulus, when compared to the third, the point enclosed in parentheses, may have occurred because listeners were able to distinguish the unmodified stimulus from a stimulus that had been modified).

The mean results of the three subjects' performance on the alignment test are shown in Figure 6. The ordinate represents the displacement from acoustic isochrony of the test stimuli relative to the reference syllable in milliseconds. The abscissa represents the amount of silence inserted into the test stimuli in milliseconds.

In this experiment P-center location, as shown by the regression line, moves linearly toward stimulus offset across the continuum with a slope of 1.00, $r = .75$. (The slopes of the P-center regression lines for the individual subjects are 1.03, .94 and 1.04; the correlations are .82, .71 and .77. Each correlation is significant at the .001 level). Thus, as in Experiment 1, there is a 1-ms shift in P-center location for every millisecond of change in gap duration. The phonetic identity of syllable-initial consonantal segments does not affect the P-center; that is, there is no abrupt shift in P-center location at the category boundary (which is indicated by the arrow). A goodness-of-fit test provided an outcome analogous to that performed on the data from Experiment 1. Thus, there is no significant departure from linearity in the data.

The results of both Experiments 1 and 2 demonstrate a millisecond-for-millisecond shift in the P-center for each stimulus manipulation. Contrary to Howell's (1984) suggestion, in this experiment the P-center varied even though the rise time of the test stimuli remained constant. We have yet to determine, however, what aspect of the changing durational pattern accounts for our results. In particular, we want to know whether P-center shifts are a function of changes in gap size, of changes in overall stimulus duration, of changes in the duration of the prevocalic segment of the syllable, or of changes in the temporal location of the acoustically defined vowel: all of these change linearly and at the same rate in Experiments 1 and 2. The remaining experiments are designed to distinguish among these alternatives.

Experiment 3

Experiment 3 was designed to control for possible influences of variation in overall syllable duration on the P-center results of Experiment 2. Stimulus duration was held constant by excising an equal amount of the

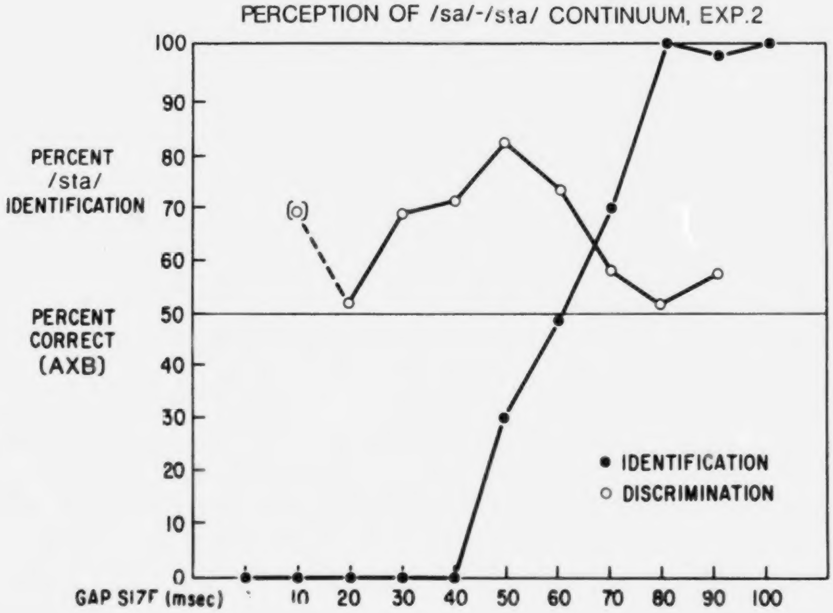


Figure 5. Mean identification and discrimination function (pooled across three subjects) for /sa/-/sta/ continuum, for Experiment 2.

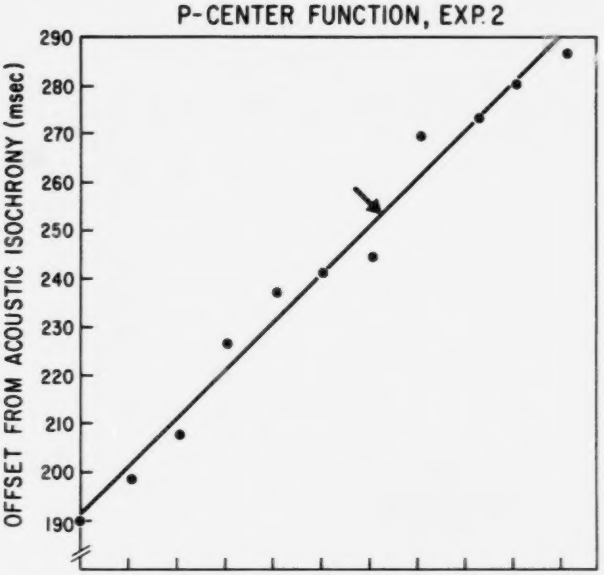


Figure 6. Mean P-center alignment function for Experiment 2, pooled across three subjects. The solid line represents the regression line and the arrow indicates the category boundary.

frication to offset the amount of silence inserted. If the shift in P-center noted in Experiment 2 was primarily related to gap duration, then the P-center functions of Experiments 2 and 3 should be equivalent. If, however, the observed shift in P-center is due either to prevocalic duration or to total duration of the syllable, then the P-center location should not change across the continuum.

Methods

Stimuli. The extreme stimuli of Experiment 3 are presented in Figure 7. The center waveform shows the unmodified /sa/ used in Experiments 2 - 4. In the final member of the continuum, represented by the upper waveform, 100 ms of silence has been inserted between the frication and the vocalic segment and 100 ms of frication has been deleted. For each stimulus in which silence was inserted, a compensatory amount of frication was excised beginning at a point 72.2 ms into the frication. This location was chosen because it allowed us both to maintain the original onset and offset characteristics of the frication and to excise a substantial amount of noise from within the fricative segment. The identification, discrimination, and alignment tests were organized as in Experiment 2.

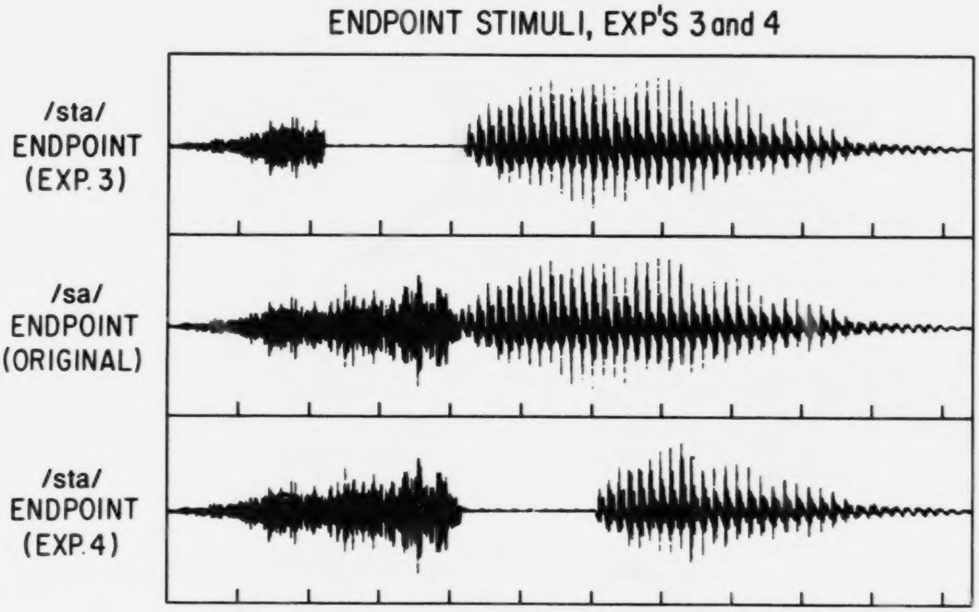


Figure 7. Continuum endpoint stimuli for Experiments 3 & 4: unmodified stimulus (center panel), extreme stimulus manipulation for Experiment 3 (upper panel), extreme stimulus manipulation for Experiment 4 (lower panel).

Subjects. The subjects were those of Experiment 2.

Results and Discussion

The identification data showed an abrupt shift in perceived phoneme category at 55 ms, and the discrimination data showed that discrimination is somewhat better near the category boundaries than within categories (Figure 8).

The alignment test for Experiment 3 (Figure 9) shows no significant change in P-center location across the continuum. The slope of the regression line is -0.003; the correlation is = 0.004. (The slopes of the P-center regression line for the individual subjects are -0.13, 0.09, and 0.02; the correlations are -0.14, 0.18, and 0.03, respectively). This result shows that the linear shift in the P-center noted in Experiment 2 is not due to the increases in gap duration, per se. Nor, consistent with Experiments 1 and 2, is P-center location affected by the phonetic identity of syllable initial prevocalic segments. And finally, this experiment also shows that P-center shifts are not necessarily affected by a syllable's amplitude envelope (as suggested by Howell, 1984), since the envelopes of the stimuli varied although the P-center did not.

The canceling of the effect of the manipulation of gap duration on P-center location may be ascribed to the canceling of its effect on total syllable duration, to the canceling of its effect on the duration of the prevocalic consonant cluster, or to the canceling of its effect on the onset time of the vocalic segment. Experiment 4 is designed to distinguish among these alternatives.

Experiment 4

In Experiment 4, syllable duration was held constant, as in Experiment 3, however, but here compensation for increases in gap duration was achieved by shortening the vocalic segment of the syllable rather than the frication. If the variation in duration of the prevocalic segment or the onset time of the vocalic segment (acoustically defined) is principally responsible for the P-center shifts noted in Experiment 2, then the P-center functions of Experiments 2 and 4 should be equivalent. If total duration is responsible for the shifts in P-center, then the P-center should remain constant across the continuum (as in Experiment 3).

Method

Stimuli. The stimulus manipulations in Experiment 4 are illustrated in Figure 7. The center waveform represents the unmodified /sa/ used in Experiments 2-4. The lower waveform represents the final stimulus used in Experiment 4, in which 12 pitch periods were excised and 93 ms of silence were inserted. In the continuum, syllable duration was held constant by excising successive pitch pulses from the vocalic segment and then inserting compensatory amounts of silence. By excising complete pitch pulses from the vowel rather than excising an arbitrary amount of the vocalic segment, we avoided abrupt discontinuities in the periodic part of the signal. Pitch pulses were extracted beginning 67.7 ms into the vocalic segment. This location was both within the steady-state portion of vowel and beyond the peak intensity of the vowel. The individual pitch pulses ranged from 7.5 to 8.0 ms in duration. The identification, discrimination, and alignment tests were organized as in Experiment 2.

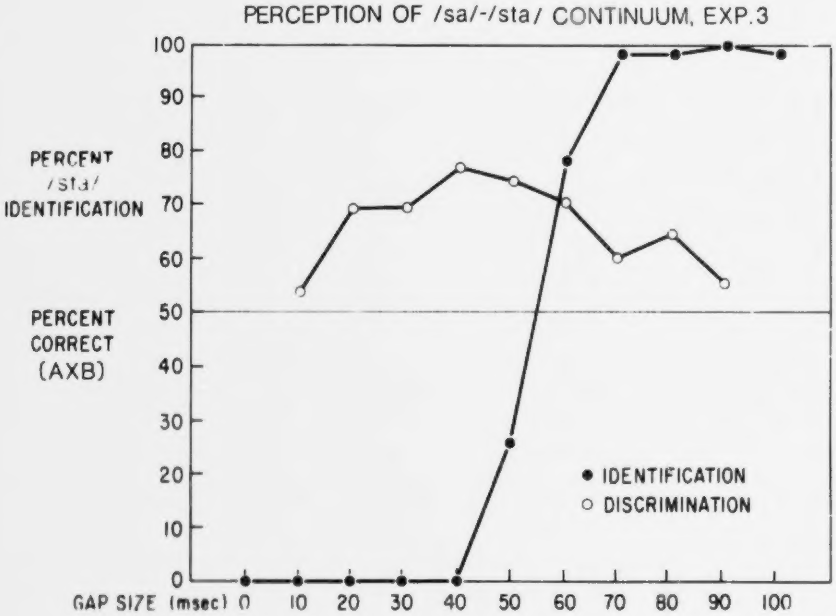


Figure 8. Mean identification and discrimination function (pooled across three subjects) for /sa/-/sta/ continuum, for Experiment 3.

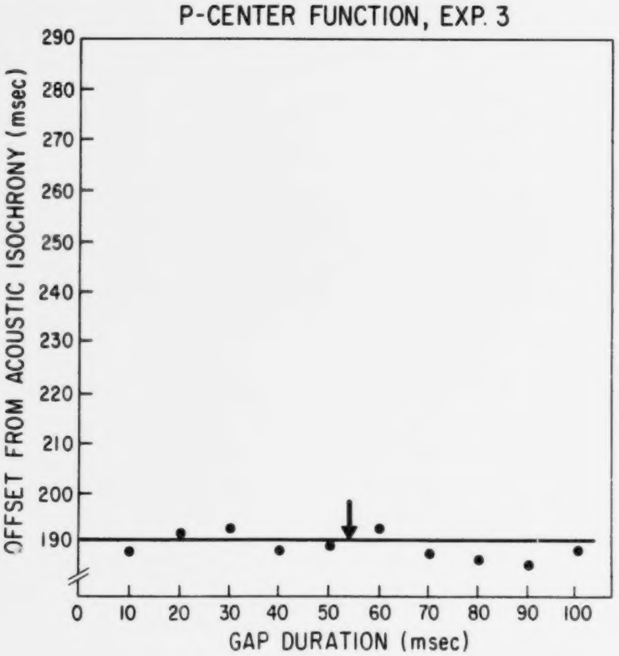


Figure 9. Mean P-center alignment function for Experiment 3, pooled across three subjects. The solid line represents the regression line and the arrow indicates the category boundary.

Subjects. The subjects were those of Experiments 2 and 3.

Results and Discussion

The identification data showed an abrupt shift in perceived phoneme category at 47 ms, and the discrimination data showed that discrimination was better near the category boundaries than within categories (Figure 10).

The results of the alignment test for Experiment 4 (Figure 11) show a linear shift in the P-center toward stimulus offset. The arrow indicates the category boundary. The slope of regression line is .83 with r = .68. (The slopes of the P-center regression lines for the individual subjects are .80, .80, and .88; the correlations are .88, .58, and .72, respectively. Each correlation is significant at the .001 level.) A dashed line with a slope of one is shown for comparison. A goodness-of-fit test reveals that the function is linear with no significant departure from linearity.

In this experiment, the P-center shifts toward stimulus offset, but the change is less than 1 ms of shift in the P-center for each millisecond of change in the gap duration. A paired t test comparing the slopes of the subjects' regression lines of Experiment 2 with those of Experiment 4 shows that these slopes are significantly different, $t(2) = 6.47$, $p < .05$. For the most extreme stimulus in Experiment 4, a gap size of 93 ms results in a P-center shift of 74.5 ms, whereas, in Experiment 2, a comparable gap size of 90 ms, shifts the P-center 91.4 ms--a difference of about 17 ms.

Our results are in agreement with those of Marcus (1981) who showed that P-center location is determined by the temporal makeup of the entire stimulus. We can interpret the results of Experiment 4 as follows. Increases in the duration of the prevocalic segment cause the P-center to shift toward stimulus offset, as it did in Experiments 1 and 2. Compensatory decreases in vowel duration, however, shift the P-center toward stimulus onset, although the magnitude of the shift is less. This interpretation is supported by Marcus (1981), who has shown that as the vocalic segment in CV syllables decreases in duration, the P-center shifts toward stimulus onset.

The results of Experiment 4 also show that P-center location cannot be simply associated with acoustically defined vowel onsets. For if the onset of the vowel were correlated with the P-center results in the previous experiments, then there should have been a millisecond shift in P-center location for every millisecond that the vowel onset shifted in the present experiment. Our results, however, clearly show that the correlation between P-center location and vowel onsets is not perfect and that P-centers cannot be linked exclusively to any single acoustically defined event in the speech signal. Other investigators have also shown that P-center location seems to be correlated with something other than a word's acoustically defined vowel onset (Allen, 1972; Fowler & Tassinary, 1981; Rapp, 1971). In Allen's study, subjects were required to tap "on the beat" of a specified syllable in a sentence. In Fowler and Tassinary's study, subjects were asked to produce rhyming nonsense syllables in time to a metronome. In Rapp's study, subjects were also asked to produce nonsense syllables in time with a regularly occurring pulse. In each of the studies, the pulse or tap both preceded the acoustic onset of the vowel and was positively correlated with the duration of the initial consonant or consonant cluster.

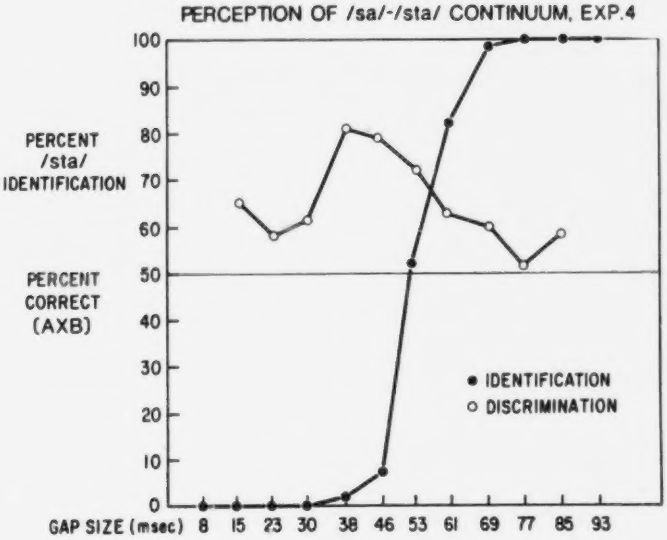


Figure 10. Mean identification and discrimination function (pooled across three subjects) for /sa/-/sta/ continuum, for Experiment 4.

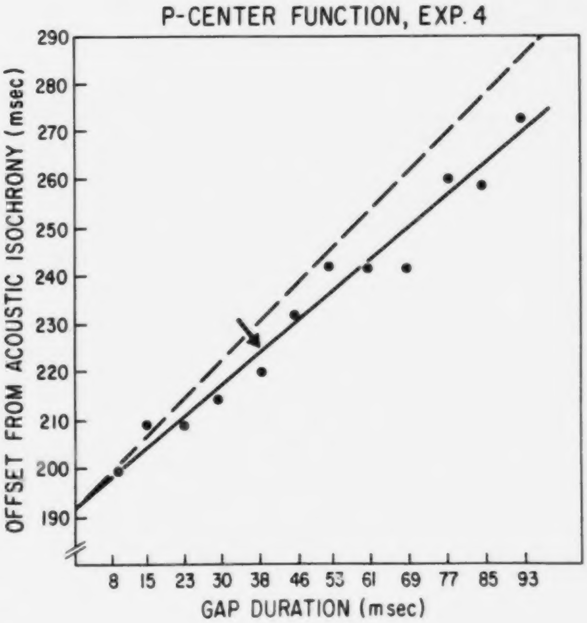


Figure 11. Mean P-center alignment function for Experiment 4, pooled across three subjects. The solid line represents the regression line, the broken line has a slope of 1 for comparison and the arrow indicates the category boundary.

Summary and General Conclusion

Our results demonstrate that a listener's judgments of the P-center of a syllable are not affected by the phonetic identity of the prevocalic segment or by any obvious acoustic properties of the signal, such as gap duration or simply the overall duration of the stimulus. Instead, the P-center appears to have been determined by a combination of at least two different aspects of the signal, the duration of the prevocalic segment and, to a lesser extent, the duration of the vocalic segment.

Our results also bear upon rationales for P-center shifts based on the amplitude envelope of speech stimuli. Howell (1984) performed experiments in which the onset intensity envelope of a /ša/ syllable was altered. He claimed that this manipulation was a "sufficient" source of variation in P-center location. Furthermore, he suggested that all of Marcus's manipulations (Marcus, 1981) in which the P-center was altered could be attributed to an alteration of the distribution of energy in the amplitude envelope. Although temporal judgments of nonspeech stimuli are influenced by their rise time characteristics (Howell, 1984; Vos & Rasch, 1981), this explanation does not account for P-center shifts in speech stimuli. For example, when the onset envelope of the test stimuli remained unaltered (Experiments 2 and 4), the P-center location varied; in contrast, when the onset envelopes of the test stimuli were altered, (Experiment 3), the P-center did not vary in location. Consistent with this finding, Tuller and Fowler (1980) radically changed the amplitude envelope of speech syllables by infinite peak clipping and found no shift in the P-center. Finally, Marcus (1981) found that increases in the silent interval for the dental stop in the word "eight" shifted the P-center toward the acoustic offset of the word, but that increases in the amplitude of the release burst did not affect the location of the P-center.

Although our findings show that the phonetic identity of syllable-initial consonants does not affect the location of the P-center, they do not rule out any possible effect of the phonetic structure of a syllable on P-center location. Given that P-center location shows a precise millisecond-for-millisecond relationship with the duration of prevocalic segments in a syllable, but is affected to a markedly smaller extent by the duration of the vocalic segment, the results of Experiment 3 are open to both phonetic and acoustic interpretations. One interpretation is that this diminished durational effect on P-center location occurs abruptly just at the point in a syllable where the listener's perception of prevocalic segments gives way to perception of the vowel. We would identify this effect as phonetic and, accordingly, we would predict that manipulations of vowel duration in vowel-initial syllables would have effects on P-center location equivalent to the vocalic effects in Experiment 4. An alternative explanation is that the effects of changes in duration are weaker the farther away they are from the syllable onset. We would identify such an effect as acoustic, and would expect the effects of durational manipulations of vowels to be greater in vowel-initial syllables than in those with initial consonants. We are currently investigating whether the phonetic quality of syllable constituents as well as the serial position of those constituents within a syllable affect the location of the P-center.

References

Allen, G. (1972). The location of rhythmic stress beats in English: An experimental study, I. Language and Speech, 15, 72-100.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 29, 191-211.

Fowler, C. A., & Tassinary, L. G. (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), Attention and Performance (Vol IX, pp. 521-535). Hillsdale, NJ: Lawrence Erlbaum Associates.

Healy, A. F., & Repp, B. H. (1982). Context sensitivity and phonetic mediation in categorical perception. Journal of Experimental Psychology: Human Perception and Performance, 8, 68-80.

Howell, P. (1984). An acoustic determinant of perceived and produced anisochrony. In M. P. R. Van den Broecke & A. Cohen (Eds.), Proceedings of the 10th International Congress of Phonetic Sciences (pp. 429-433). Dordrecht: Foris Publications.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Language and Speech, 4, 175-195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 54, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. Journal of Experimental Psychology, 61, 379-388.

Marcus, S. M. (1976). Perceptual centres. Unpublished doctoral dissertation, Cambridge University.

Marcus, S. (1981). Acoustic determinants of perceptual center (P-center) location. Perception & Psychophysics, 30, 247-256.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). Psychological Review, 83, 405-408.

Rapp, K. (1971). A study of syllable-timing (Papers in Linguistics). Stockholm: University of Stockholm.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), Speech and language: Advances in basic research and practice (Vol. 10). Orlando, FL: Academic Press.

Tuller, B., & Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. Perception & Psychophysics, 27, 277-283.

Vos, J., & Rasch, R. (1981). The perceptual onset of musical tones. Perception & Psychophysics, 29, 323-335. 68-80.

TWO CHEERS FOR DIRECT REALISM*

Michael Studdert-Kennedy†

> "Beware Procrustes, bearing Occam's razor."
> -- Lise Menn

I am very much in sympathy with Fowler's approach (henceforth, CAF) because it is grounded in a functionalist, biological view of language. No doubt the approach will be faulted, despite its disclaimers, for narrowly focussing on phonetic structure. Yet what is new in CAF is precisely its scope: the range of phonetic fact for which it takes responsibility. Basic research in speech perception and basic research in speech production (no less than applied research in speech synthesis and machine recognition) have tended to follow parallel lines. Perceptual research typically manipulates acoustic variables with little regard for articulatory constraints, while production research typically studies the actions of individual muscles or articulators with little concern for how they are coordinated to yield a perceptually coherent acoustic signal. By adopting a single abstract unit (corresponding to the phoneme-sized phonetic segment) as the presumed functional element of both production and perception, CAF lays the ground for a program of research responsible to both. Nor is it coincidence that the selected unit is potentially alphabetic. For CAF thus acknowledges that our accounts of speaking and listening must be consistent with the facts of writing and reading.

A signal virtue of CAF, then, is that it accepts responsibility for the segmental structure of all four modes of language action: like any good theory, it proposes to unify (eventually) related classes of fact that are commonly treated as separate. The faults of CAF largely stem, I believe, from a somewhat too zealous attempt to impose a framework, devised to handle an animal's traffic with the physical world, on a communication system with a quite different evolutionary history and function.

CAF includes three assumptions that need to be modified or, at least, explicated: (1) perception is "unmediated by cognitive processes of inferencing or hypothesis testing"; (2) listeners "extract information about

---

*Journal of Phonetics, 1986, 14, 99-104. Commentary on Fowler, C. A.: An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, 1986, 14, 3-28. Also SR-85, this volume.

†Also Queens College and Graduate Center, City University of New York

Acknowledgment. This comment was written while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, CA. My thanks to the Spencer Foundation for financial support, and to Björn Lindblom and Peter MacNeilage for discussion and comments.

articulation from the acoustic speech signal"; (3) "it matters little through what sense we realize what speech event has occurred." My comments follow.

Unmediated Perception

A corollary of this assumption seems to be that the phonetic segment should not be constructed, in either perception or production, from smaller units. Accordingly CAF, invoking speech error data to support the choice of unit, implicitly dismisses "feature" errors as unimportant. Yet such errors do occur, with some low frequency, and have to be accounted for. Voicing metathesis seems to be the most common (e.g., *clear blue sky* → *glear plue sky* (Fromkin, 1971)), but place metathesis also occurs (e.g., *pedestrian* → *tebestrian* (Fromkin, 1971); *wild goose chase* → *wild juice case* (Robert Remez, personal communication). These errors are interesting because they reflect a level of organization below the segment.

The possibility of such errors is implicit in CAF's definition of a phonetic segment as a "set of coordinated gestures." Elsewhere, Fowler and her colleagues (Fowler, Rubin, Remez, & Turvey, 1980) treat the phonetic segment as a set of nested, or embedded, coordinative structures that arise as functional groupings of muscles, marshalled for moment-to-moment control of speech. The coordinative structures of Fowler et al. evidently correspond to the gestures of CAF. Similarly, Kelso, Saltzman, and Tuller (1986) discuss the task-specific grouping of muscles to execute a gesture, nested within the CV syllable. CAF, quite properly in my view, regards these gestures as non-linguistic (or non-phonetic): "lip closure *per se* is not an articulatory speech event." Lip closure only becomes phonetic (i.e., only performs a linguistic function) by virtue of its coordination with other non-phonetic gestures in an appropriate linguistic context.

A speaker, then, is engaged in moment-to-moment marshalling of intrinsically functionless muscle systems to fulfill a phonetic function--much as a tennis player marshalls muscles to execute a tennis stroke. A skilled speaker has a repertoire of routinized processes that assemble non-phonetic gestures into phonetic segments. Errors in gestural assemblage may then be rare because the process occurs with very high frequency, so that a given gesture is called into a phonetic segment even more frequently than a phonetic segment is called into a syllable. Errors in the process may also be rare due to tight anatomical and physiological constraints on gestural coupling: Voicing metathesis is perhaps the most common error because voicing is relatively loosely coupled to supralaryngeal action. In any event, by this account, a gestural error is motoric, a segmental error phonetic.

Consider now the child learning to speak. Its task is to discover how to marshall its repertoire of non-linguistic babbling gestures for linguistic use. Its first linguistic (functionally communicative) segments are words or formulaic phrases. The child evidently perceives these units as constructed from non-linguistic gestures. For example, Ferguson and Farwell (1975) report the following attempts by a 15-month-old child to say the word *pen*:

[mã^{ə}, ˇʌ̃, dɛ$^{dn̩}$, hɪn, ^{m}bõ, p^{h}ɪn, t^{h}n̩t^{h}n̩t^{h}n̩, ba^{h}, d̥^{h}au^{N}, buã].

In these attempts, we find all the gestures required to utter *pen*: lip closure, lingua-alveolar closure, tongue raising and fronting, velum raising and lowering, glottal narrowing and spreading. The gestures are misordered and mistimed, but it is evident that the acoustic structure of the word did specify for the child the gestures that compose it.

As the child develops, it will come to recognize recurrent gestural groupings as functional elements in speaking: the phonetic segment will emerge as the interface between non-linguistic gesture and linguistic word. Will the child thereby lose its capacity to perceive gestures? It would seem not. The speech error data demonstrate that the adult may produce gestures separately from the segmental structure in which they are normally embedded. If the perspectives of speakers and listeners are "interchangeable," as CAF proposes, listeners must assemble segments from non-phonetic, auditory markers in the signal no less than speakers assemble them from a non-phonetic gestural repertoire. This may not call for "inferencing or hypothesis testing" in perception, but it does call for some process less immediate than the word "direct" would seem to imply.

Extracting Information about Articulation

Direct realism presses CAF into "defining speech event interchangeably from the perspectives of talkers and listeners." For the definition to hold we must assume that the problem of functional equivalence among diverse motor patterns, in general, or of the many-to-one relation between articulation and acoustics, in particular, has been solved (cf. Kelso, et al., 1986/this volume). We could then be confident that articulation and acoustics are, at some abstract level of description, fully isomorphic: to every acoustic pattern of change in frequency and time there exactly corresponds an articulatory pattern of movement in space and time, and vice versa.

Ironically, this assumption renders ambiguous much of the evidence cited to support it. To show that listeners extract information about articulation from the speech signal, CAF cites several studies in which listeners' perceptual judgments seemed to be in better agreement with the articulatory pattern than with the acoustic. Such findings are anomalous, if articulatory and acoustic patterns are isomorphic. For the "P-center" studies CAF resolves the anomaly by arguing that it arose from an error in the conventional acoustic measurements of vowel onset. Once the error was corrected, acoustics, articulation, and perception fell into line.

An equivalent move in the /slIt-/splIt/ "trading relations" phenomenon would require systematic measurement of the articulatory correlates of acoustic silent interval (stop closure) and formant transitions (stop release). Such measurements have never been reported, so far as I know, and in the cited study they could not be appropriately made because the experiments were done with synthetic speech. Articulatory equivalence (or non-equivalence) was therefore inferred, with some circularity, from perceptual equivalence (or non-equivalence). However, if the appropriate measurements were done on natural speech, articulation, acoustics, and perception would, by the hypothesis of CAF, again fall into line.

In short, if acoustics and articulation are fully isomorphic, they are merely notational variants. Whether we describe the listener as perceiving sound patterns or as perceiving articulatory patterns, is then a matter of theoretical taste. Direct perception of articulation becomes merely an axiom of a direct-realist theory.

Perhaps all this is sophistry. We know, after all, that listeners do extract information about articulation. How otherwise would every normal child come to speak the dialect of its peers? We know too from studies of "lip-reading" that acoustic and optic information about speech may combine in perception. These studies suggest that we are able to imitate or repeat the utterance of another because perception extracts an amodal pattern of information, isomorphic with the pattern that controls articulation--just as CAF claims. What seems to be at issue then is not whether listeners can extract information about articulation, but whether they always do, and whether perception is direct, in the sense that the medium structured by articulation is transparent and a matter of indifference to the perceiver.

The Medium of Amodality

Each species of animal has a unique combination of perceptual and motoric capacities. Characteristic motor systems have evolved for locomotion, predation, consumption, and mating. Matching perceptual systems have evolved to guide the animal in these activities. The selection pressures shaping each species' perceptuomotor capacities have come, in the first instance, from physical properties of the world.

By contrast, these perceptuomotor capacities themselves must have played a crucial role in shaping the form of a social species' communication system. The general point was made by Huxley (1914) when he remarked that the elaborate courtship rituals of the great crested grebe must have evolved by selection of perceptually salient patterns from the bird's repertoire of motorically possible actions. Certainly, specialized neuroanatomical signaling devices have often evolved, but they have typically done so by modifying pre-existing structures just enough for them to perform their new function without appreciable loss of their old. The cricket stridulates with its wings, the grasshopper with its legs; birds and mammals vocalize with their eating and breathing apparatus. The quality and range of possible signals is thus limited by the structure and function of the co-opted mechanism.

A further constraint on signal form must come from the perceptual system to which the signals are addressed. Here again specialized devices (e.g., feature detecting systems, templates) have certainly evolved, presumably by some minimal modification of a pre-existing perceptual system. Typically, such specialized devices, in the auditory realm, seem to have evolved in animals with little or no parental care and therefore little opportunity to learn their species' call: bullfrogs, treefrogs, certain species of bird, and so on. We have no evidence for such devices in the human.

We are not then surprised that the main speech frequencies are spread over the three octaves (500-4000 Hz) to which the human auditory system is most sensitive, and that (as the quality of deaf speech attests) speech sounds have evolved to be heard, not seen. Thus, the differences in degree of constriction among high vowels, intra-oral fricatives, and stops are highly salient auditorily; but the same differences in, say, finger to thumb distance, would be scarcely detectable if they were incorporated in a visual sign language. Similarly, the abrupt acoustic changes at the onset of many CV syllables may have been favored, in part, because the mammalian auditory system is particularly sensitive to such discontinuities (Delgutte, 1982; Kiang, 1980; Stevens, 1981). The resulting auditory contrast perhaps facilitates the listener's perceptual segmentation both of the syllable from its context and of the consonant from its following vowel.

On the other hand, the signs of American Sign Language have evolved (over the past 170 years) to be seen, not heard. Accordingly, signs formed at the center of the signing space (that is, in the foveal region of the viewer) tend to use smaller movements and smaller handshape contrasts than signs formed at the periphery (Siple, 1978).

In short, even if the sense that informs us about our environment "matters little" in the farmyard (itself a dubious claim), it seems not to "matter little" for communication. Language has evolved within the constraints of pre-existing perceptual and motor systems. We surrender much of our power to understand that evolution, if we disregard the properties of those systems. And indeed, CAF concedes as much by citing with approval Lindblom's work on the emergence of phonetic structure. The success of that work, particularly for vowel systems, rests on an acoustic description of speech sounds, weighted according to a model of the auditory system, and on the use of an auditory distance metric to assess their perceptual distinctiveness.

How, then, are we to square the auditory properties of the speech signal with the evident amodality of the speech percept? We must, I think, question CAF's definition of speech events as "a talker's phonetically structured articulations." A speech event is not simply articulation, however structured, any more than a tennis serve is simply the server's swing. A speech event, even narrowly conceived as phonetic action, only occurs when a speaker executes, and a listener apprehends, a phonetic function. Elsewhere, Fowler (1980) has termed this function the talker's phonetic "intent" (cf. Liberman, 1982). "Intent" seems to correspond, at least in level of abstraction, to task (or goal), the level at which Kelso et al. (1986/this volume) define a single function from which different, but equivalent, articulations may arise. Surely, this too must be the level--free of adventitious articulatory variation and its acoustic consequences--at which the listener's percept might properly be termed amodal.

Looked at in this way, articulation becomes as much a medium of speech, structured by the talker's goals, as the acoustic signal, structured by the talker's articulations, and as its heard counterpart, structured by the listener's (suitably "attuned") auditory system. Each medium is then subject to its own characteristic type of variability.

One happy side-effect of setting speaker and listener (articulation and audition) on equal footing is that we can rationalize perceptual error more simply than does CAF. The likelihood of an error is a function of its cost. Collisions between swallows, swarming in hundreds through a cloud of insects, or between pelicans flocking and diving into a school of fish, are rare (though, *pace* direct realism, they do occur!). Natural selection prunes the error-prone from the species, honing the perceptuomotor systems of the survivors to a fine precision. By contrast, errors in speaking and listening carry essentially no penalty. Moreover, if phonetic form has been shaped by compromise between the articulatory capacities of a speaker and the perceptual capacities of a listener, we might expect some instability in phonetic execution, some slight oscillation between the opacity comfortable for a speaker, the transparency called for by a listener (cf. Slobin, 1980). We may view a conversation as a microcosm of evolution: the speaker balances a desire to be understood against articulatory ease, the listener a desire to understand against the costs of attention (Lindblom, 1983). Given these conflicting demands and the modest penalties for error, we might even be surprised that errors are not more frequent than they are. In this regard, while no one, so far as I know, has studied the social contexts in which perceptual errors occur, they are probably rare when the speaker is, say, delivering instructions for a parachute jump.

In conclusion, the fact that we hear speech is no less important and no more accidental than the fact that we articulate it. Many of the long-standing problems of speech research, including normalization, segmentation, and even the lack of invariance, may be illuminated by an understanding of audition. Even if the information we extract is amodal, just what information we extract and the precision with which we extract it depend on our auditory sensitivity.

References

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of the auditory nerve. In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system* (pp. 131-149). New York: Elsevier.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition: English initial consonants in the first fifty words. *Language*, *51*, 419-430.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*, 113-133.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 373-420). New York: Academic Press.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27-52.

Huxley, J. S. (1914). The courtship habits of the great crested grebe (*Podiceps cristatus*), with an addition on the theory of sexual selection. *Proceedings of Zoological Society* (London), *XXXV*, 491-562.

Kelso, J. A.S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*, 29-59. Also *SR-85*, this volume.

Kiang, N. Y. S. (1980). Processing of speech by the auditory nervous system. *Journal of the Acoustical Society of America*, *68*, 830-835.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist*, *37*, 148-167.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217-245). New York: Springer-Verlag.

Siple, P. (1978). Visual constraints for sign language communication. *Sign Language Studies*, *19*, 97-112.

Slobin, D. I. (1980). The repeated path between transparency and opacity in language. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 229-243). Deerfield Beach, FL: Verlag Chemie.

Stevens, K. N. (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics and psychoacoustics. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 61-74). New York: North-Holland.

AN EVENT APPROACH TO THE STUDY OF SPEECH PERCEPTION FROM A DIRECT-REALIST PERSPECTIVE*

Carol A. Fowler†

1. Introduction

There is, as yet, no developed event approach to a theory of speech perception and, accordingly, no body of research designed from that theoretical perspective. I will offer my view as to the form that the theory will take, citing relevant research findings where they are available. The theory places constraints on a theory of speech production, too. Therefore, I will also have something to say about how talkers must talk for an event approach to be tenable. I will begin by defining the domain of the theory as I will consider it here.

An ecological event is an occurrence in the environment defined with respect to potential participants in it. Like most ecological events (henceforth, events), one in which linguistic communication takes place is highly structured and complex. Accordingly, it can be decomposed for study in many different ways. One way in which it is almost invariably decomposed by psycholinguists and linguists is into the linguistic utterance itself on the one hand and everything else on the other. In ordinary settings in which communication takes place, this is almost certainly not a natural partitioning because it leaves out several aspects of the setting that contribute interactively with the linguistic utterance itself to the communication. These include the talker's gestures (McNeill, 1985), aspects of the environment that allow the talker to point rather than to refer verbally, and the audience whose shared experiences with the talker affect his or her speaking style. The consequences of making this cut have not been worked out, but, at least for purposes of studying language as communication, they may be substantial (cf. Beattie, 1983). For the present, however, I will preserve the partitioning and one within that as well.

The linguist, Hockett (1960), points out that languages have "duality of patterning"--that is, they have words organized grammatically into sentences, and phonetic segments organized phonotactically into words. Both levels are essential to the communicative power of language.

Grammatical organization of words into sentences gives linguistic utterances two kinds of power. First, the communicative content of an

---

*Journal of Phonetics, 1986, 14, 3-28.

†Also Dartmouth College

Acknowledgment. Preparation of this paper was supported by NICHD Grant HD 16591 to Haskins Laboratories. I thank Ignatius Mattingly for his comments on an earlier draft of the manuscript.

utterance is superadditive with respect to the contents of the words composing the sentences taken as individuals. Second, talkers can produce novel utterances that the audience has not heard before; yet the utterance can convey the talkers' message to the audience. I will refer to a linguistic utterance at this level of description as a "linguistic event," and, having defined it, I will have little else to say about it until the final section of the paper.

The second structural tier, in which phonetic segments constitute words, supports an indefinitely large lexicon. Were each word to consist of a holistic articulatory gesture rather than a phonotactically-organized sequence of phonetic segments, our lexicons would be severely limited in size. Indeed, recent simulations by Lindblom (Lindblom, MacNeilage, & Studdert-Kennedy, 1983) show that, as the size of the lexicon is increased (under certain constraints on how new word labels are selected), phonetic structure emerges almost inevitably from a lexicon consisting initially of holistic closing and opening gestures of the vocal tract. These simulations may show how and why phonetic structure emerged in the evolution of spoken language and how and why it emerges in ontogeny.

I will refer to a talker's phonetically-structured articulations as "speech events." It is the perception of these events that constitutes the major topic of the paper. A speech event may also be defined as a linguistic utterance having phonetic structure as perceived by a listener. In defining a speech event interchangeably from the perspectives of talkers and listeners, I am making the claim, following others (e.g., Shaw, Turvey, & Mace, 1982) that a theory of event perception will adopt a "direct realist" stance. According to Shaw et al.:

> Some form of realism must be captured in any theory that claims to be a theory of perception. To do otherwise would render impossible an explanation of the practical success of perceptually guided activity. (p. 159)

That is, to explain the success of perceptually-guided activity, perception is assumed to recover events in the real world. For this to be possible consistently (see Shaw & Bransford, 1977), perception must be direct, and in particular unmediated by cognitive processes of inferencing or hypothesis testing, which introduce the possibility of error.[1]

By focusing largely on speech events, I will be discussing speech at a level at which it consists of phonetically-structured syllables, but not, necessarily of grammatical, meaningful utterances. It is ironic, perhaps, that a presentation at a conference on event perception should focus on a linguistic level that is not transparently significant ecologically. However, speech events can be defended as natural partitionings of linguistic events--that is, they can be defended as ecological events--and there is important work to be done by event theorists even here.

The defense is that talkers produce phonetically-structured speech, listeners perceive it as such, and they use the phonetic structure they perceive to guide their subsequent behavior. Talkers reveal that they produce phonetically-structured words when they make speech errors. Most submorphemic errors are misorderings or substitutions of single phonetic segments (e.g., Shattuck-Hufnagel, 1983). For their part, listeners can be shown to extract phonetic structure from a speech communication at least in certain experimental settings. That they extract it generally, however, is suggested by the observation that they use phonetic variation to mark their identification with a social group or to adjust their speaking style to the conversational setting. Of course, infant perceivers must recover phonetic structure if they are to become talkers who make segmental speech errors.

This defense is not intended to suggest that the study of perception of speech events is primary or privileged in any sense. It is only to defend it as one of the partitionings of an event involving linguistic communication that is perceived and used by listeners; therefore is an event in its own right and requires explanation by a theory of perception.

I will discuss an event approach to phonetic perception in the next three major sections of the paper. The first two sections consider direct perception, first of local, short-term events, and next of longer ones. The third section considers some affordances of phonetically-structured speech.

Although there is lots of work to be done at this more fine-grained of the dual levels of structure in language, there are also great challenges to an event theory offered by language considered as syntactically-structured words that convey a message to a listener. I will discuss just two of these challenges briefly at the end of the paper and will suggest a perspective on linguistic events that an event theory might take.

2. Perception of Speech Events: A Local Perspective

There is a general paradigm that all instances of perception appear to fit. Perception requires events in the environment ("distal events"), and one or more "informational media"--that is, sources of information about distal events in energy media that can stimulate the sense organs--and a perceiver. As already noted, objects and occurrences in the environment are generally capable of multiple descriptions. Those that are relevant to a perceiver refer to "distal events." They have "affordances"--that is, sets of possibilities for interaction with them by the perceiver. (Affordances are "what [things] furnish, for good or ill" [Gibson, 1967/1982; see also, Gibson, 1979]). An informational medium, including reflected light, acoustic signals, and the perceiver's own skin, acquires structure from an environmental event specific to certain properties of the event; because it acquires structure in this way, the medium can provide information about the event properties to a sensitive perceiver. A second crucial characteristic of an informational medium is that it can convey its information to perceivers by stimulating their sense organs and imparting some of its structure to them. By virtue of these two characteristics, informational media enable direct perception of environmental events. The final ingredient in the paradigm is a perceiver who actively seeks out information relevant to his or her current needs or concerns. Perceivers are active in two senses. They move around in the environment to intercept relevant sources of information. In addition, in ways not yet well understood, they "attune" their perceptual systems (e.g., Gibson, 1966/1982) to attend selectively to different aspects of available environmental structure.

In speech perception, the distal event considered locally is the articulating vocal tract. How it is best described to reflect its psychologically significant properties is a problem for investigators of speech perception as well as of speech production. However, I will only characterize articulation in general terms here, leaving its more precise description to Kelso, Saltzman and Tuller (1986/this volume) in their presentation. One thing we do know is that phonetic segments are realized as coordinated gestures of vocal-tract structures--that is, as coupled relationships among structures that jointly realize the segments (e.g., Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984). Therefore, studies of the activities of individual muscles or even individual articulators will not reveal the systems that constitute articulated phonetic segments.

The acoustic speech signal has the characteristics of an informational medium. It acquires structure from the activities of the vocal tract and it can impart its structure to an auditory perceptual system, thereby conveying its information to a sensitive perceiver. In this way, it enables direct perception of the environmental source of its structure, the activities of the vocal tract. Having perceived an utterance, a listener has perceived the various "affordances" of the conversational event and can guide his or her subsequent activities accordingly.

This, in outline form, is a theory of the direct perception of speech events. The theory promotes a research program having four parts, three relating to the conditions supporting direct perception of speech events and the last relating to the work that speech events do in the environment. To assess the claim that speech events are directly perceived, the articulatory realizations of phonetic segments must be uncovered and their acoustic consequences identified. Next, the listener's sensitivity to, and use of, the acoustic information must be pinned down. Finally, the listener's use of the structure in guiding his or her activities must be studied. Although, of course, a great deal of research has been done on articulation and perception of speech, very little has been conducted from the theoretical perspective of an event theory and very little falls within the research program just outlined.

Indeed, my impression, based on publishing investigations of speech conducted from this perspective and on presentations of the theoretical perspective to other speech researchers, is that it has substantial face invalidity. There are several things seemingly true of speech production and perception that, in the view of many speech researchers, preclude development of a theory of direct perception of speech events. I will consider four barriers to the theory, and along with some suggestions concerning ways to surmount or circumvent them.

2.1 The First Barrier: If Listener's Recover Articulation, Why Don't They Know It?

A claim that perceivers see environmental events rather than the optic array that stimulates their visual systems seems far less radical than a claim that they hear phonetically-structured articulatory gestures rather than the acoustic speech signal. Indeed, when Repp (1981) makes the argument that phonetic segments are "abstractions" and products of cognitive processes applied to stimulation, he says of them that "*they have no physical properties*--such as duration, spectrum and amplitude--and, therefore, *cannot be measured* (p. 1463, italics in the original). That is, he assumes that if phonetic segments were to have physical properties, the properties would be acoustic. Yet no one thinks that, if the objects of visual perception--that is, trees, tables, people, etc--do have physical properties, their properties are those of reflected light.

Somewhat compatibly, our phenomenal experience when we hear speech certainly is not of lips closing, jaws raising, velums lowering, and so on, although our visual experience is of the objects and events in the world. Of course, we do not experience surface features of the acoustic signal either--that is, silent gaps followed by stop-bursts, or formant patterns, or nasal resonances.

I cannot explain the failure of our intuitions in speech to recognize that perceived phonetic events are articulatory, as compared to our intuitions about vision, which do recognize that perceived events are environmental, but I can think of a circumstance that exacerbates the failure among researchers. If, in an experimental study, listeners do indeed recover articulatory events in perception, there is likely to be a large mismatch between the level of description of an articulatory event that they recover in an experimental study and a researcher's description of the activities of the individual articulators. That is, speech researchers do not yet know what articulatory events consist of. If a perceiver does not experience "lips closing," for example, that is as it should be, because lip closure *per se* is not an articulatory speech event. Rather (see the contribution by Kelso et al.), an articulatory event that is a phonetic event, for example, is a coordinated set of movements by vocal tract structures. By hypothesis, the percept [b] corresponds to extraction from the acoustic speech signal of information that the appropriate coordinated gestures occurred in the talker's vocal tract--just as as the perceptual experience of a zooming baseball corresponds to extraction of information from the optic array that the event of zooming occurred in the environment.

The literature offers evidence from a wide variety of sources that listeners do extract information about articulation from the acoustic speech signal. Much of this evidence has recently been reviewed by Liberman and Mattingly (1985) in support of a motor theory.[2] I will select just a few examples.

1. *Perceptual equivalence of distinct acoustic "cues" specifying the same articulatory event*. In nonphonetic contexts, silence produces a very different perceptual experience from a set of formant transitions. However, interposed between frication for an [s] and a syllable sounding like [lIt] in isolation, they may not (Fitch, Halwes, Erickson, & Liberman, 1980). An appropriate interval of silence may foster perception of [p]; so may a lesser amount of silence, insufficient to cue a [p] percept in itself, followed by transitions characteristic of [p] release. Strikingly, a pair of syllables differing both in the duration of silence after the the [s] frication and in presence or absence of [p] transitions following the silence are either highly discriminable--and more discriminable than a pair of syllables differing along just one of these dimensions--or nearly indiscriminable--and *less* discriminable than a pair differing in just one dimension--depending on whether the silence and transitions "cooperate" or "conflict." They cooperate if, within one syllable, both acoustic segments provide evidence for stop production and, within the other, they do not. They conflict if the syllable having a relatively long interval of silence appropriate to stop closure lacks the formant transitions characteristic of stop release, while the syllable with a short interval of silence has transitions. Depending on the durations

of silence, these latter syllables may both sound like "split" or both like "slit."

The important point is that very different acoustic properties sound similar or the same just when the information they convey about articulation is similar or the same. It should follow, and does, that when an articulation causes a variety of acoustic effects (for example, Lisker, [1978], has identified more than a dozen distinctions between voiced and voiceless stops intervocalically), the acoustic consequences individually tend to be sufficient to give rise to the appropriate perception but none are necessary. (See Liberman & Mattingly, 1985, for a review of those findings.)

2. *Different perceptual experiences of the same acoustic segment just when it specifies different distal sources.* By the same token, the same acoustic segment in different contexts, where it specifies different articulations or none at all, sounds quite different to perceivers. In the experiment by Fitch et al. just described, a set of transitions characteristic of release of a bilabial stop will only give rise to a stop percept in that context if it is preceded by sufficient sufficient silence. This cannot be because, in the absence of silence, the [s] frication masks the transitions; other research demonstrates that transitions at fricative release themselves do contribute to fricative place perception (e.g., Harris, 1958; Whalen, 1981).Rather, it seems, release can only be perceived in this context given sufficient evidence for prior stop closure. Similarly, if transitions are presented in isolation where, of course, they do not signal stop release, or even production by a vocal tract at all, they sound more-or-less the way that they look on a visual display--that is, like frequency rises and falls (e.g., Mattingly, Liberman, Syrdal, & Halwes, 1971).

3. "*P center*." Spoken digits (Morton, Marcus, & Frankish, 1976) or nonsense monosyllables (Fowler, 1979), aligned so that their onsets of acoustic energy are isochronous, do not sound isochronous to listeners. Asked to adjust the timing of pairs of digits (Marcus, 1981) or monosyllables (Cooper, Whalen, & Fowler, 1984) produced repeatedly in alternation so that they sound isochronous, listeners introduce systematic departures from measured isochrony--just those that talkers introduce if they produce the same utterances to a real (Fowler & Tassinary, 1981; Rapp, 1971) or imaginary (Fowler, 1979; Tuller & Fowler, 1980) metronome. Measures of muscular activity supporting the talkers' articulations is isochronous in rhyming monosyllables produced to an imaginary metronome. Thus, talkers follow instructions to produce isochronous sequences, but due (in large part) to the different times after articulatory onset that different phonetic segments have their onsets of acoustic energy, acoustic measurements of their productions suggest a failure of isochrony. For their part, listeners appear to hear through the speech signal to the timing of the articulations.

4. *Lip reading*. Liberman and Mattingly (1985) describe a study in which an acoustic signal for a production of [ba] synchronized to a face mouthing [bɛ], [vɛ] and [ðɛ] may be heard as [ba], [va] and [ða], respectively (cf. McGurk & MacDonald, 1976). Listeners experience hearing syllables with properties that are composites of what is seen and heard, and they have no sense that place information is acquired largely visually and vowel information auditorily. (This is reminiscent of the quotation from Hornbostel [1927] reprinted in Gibson [1966]: "it matters little through which sense I realize that in the dark I have blundered into a pigsty." Likewise, it seems, it matters little through what sense we realize what speech event has

occurred.) Within limits anyway, information about articulation gives rise to an experience of hearing speech, whether the information is in the optic array or in the acoustic signal.

2.2 The Second Barrier: Linguistic Units Are Not Literally Articulated

A theory of perception of speech events is disconfirmed if the linguistic constituents of communications between talkers and listeners do not make public appearances. There are two kinds of reason for doubting that they do, both relating to an incommensurability that many theorists and researchers have identified between knowing and doing, between competence and performance, or even between the mental and the physical realizations of language.

One kind of incommensurability is graphically illustrated by Hockett's Easter egg analogy (Hockett, 1955). According to the analogy, articulation, and, in particular the coarticulation that inertial and other physical properties of the vocal tract requires, obliterates the discrete, context-free phonetic segments of the talker's planned linguistic message. Hockett suggests that the articulation of planned phonetic segments is analogous to the effects that a wringer would have on an array of (raw) Easter eggs. If the analogy is apt, and listeners nonetheless can recover the phonetic segments of the talker's plan, then direct detection of articulatory gestures in perception cannot fully explain perception, because the gestures themselves provide a distorted representation of the segments. To explain recovery of phonetic segments from the necessarily impoverished information in the acoustic signal, reconstructive processes or other processes involving cognitive mediation (Hammarberg, 1976, 1982; Hockett, 1955; Neisser, 1967; Repp, 1981) or noncognitive mediation (Liberman & Mattingly, 1985) must be invoked.

Hockett is not the only theorist to propose that ideal phonetic segments are distorted by the vocal tract. For example, MacNeilage and Ladefoged describe planned segments as discrete, static, and context-free, whereas uttered segments are overlapped, dynamic, and context-sensitive.

A related view expressed by several researchers is that linguistic units are mental things that, thereby, cannot be identified with any set of articulatory or acoustic characteristics. For example:

> [Phonetic segments] are abstractions. They are the end result of complex perceptual and cognitive processes in the listener's brain. (Repp, 1981, p. 1462)
>
> They [phonetic categories] have no physical properties. (Repp, 1981, p. 1463)
>
> Segments cannot be objectively observed to exist in the speech signal nor in the flow of articulatory movements...[T]he concept of segment is brought to bear *a priori* on the study of physical-physiological aspects of language. (Hammarberg, 1976, p. 355)
>
> [T]he segment is internally generated, the creature of some kind of perceptual-cognitive process. (Hammarberg, 1976, p. 355)

This point of view, of course, requires a mentalist theory of perception.

For a realist event theory to be possible, what modifications to these views are required? The essential modification is to our conceptualization of the relation between knowing and doing. First, phonetic segments as we know them can have only properties that can be realized in articulation. Indeed, from an event perspective, the primary reality of the phonetic segment is its public realization as vocal-tract activity. What we know of the segments, we know from hearing them produced by other talkers or by producing them ourselves. Second, the idea that speech production involves a translation from a mental domain into a physical, nonmental domain such as the vocal tract must be discarded.

With respect to the first point, we can avoid the metaphor of Hockett's wringer if we can avoid somehow ascribing properties to phonetic segments that vocal tracts cannot realize. In view of the fact that phonetic segments evolved to be spoken, and indeed, that we have evolved to speak them (Lieberman, 1982), this does not seem to be a radical endeavor.

Vocal tracts cannot produce a string of static shapes, so for an event theory to be possible, phonetic segments cannot be inherently static. Likewise, vocal tracts cannot produce the segments discretely, if discrete means "nonoverlapping." However, neither of these properties is crucial to the work that phonetic segments do in a linguistic communication and therefore can be abandoned without loss.

Phonetic segments do need to be separate one from the other and serially ordered, however, and Hockett's Easter egg analogy suggests that they are not. My own reading of the literature on coarticulation, however, is that the Easter egg analogy is misleading and wrong. Figure 1 is a redrawing of a figure from Carney and Moll (1971). It is an outline drawing of the vocal tract with three tongue shapes superimposed. The shapes were obtained by cine-

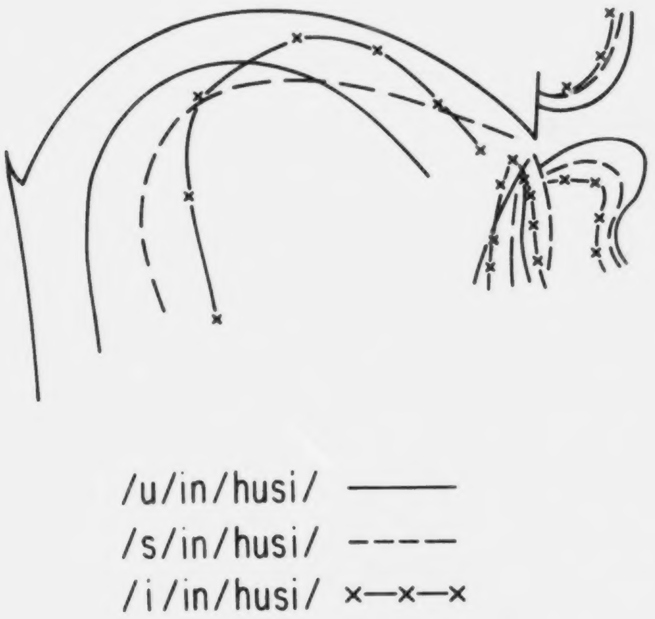


Figure 1: Cinefluorographic tracing of the vocal tract during three phases in production of /husi/ (redrawn from Carney & Moll, 1971).

fluorography at three points in time during the production of the disyllable [husi]. The solid line reflects the tongue shape during a central portion of the vowel [u]; the dashed line is the tongue shape during closure for [s]; the x-ed line is the tongue shape during a central portion of [i]. Thus, the figure shows a smooth vowel-to vowel gesture of the tongue body taking place during closure of [s] (cf. Öhman, 1966). The picture these data reveal is much cleaner than the Easter egg metaphor would suggest. Gestures for different segments overlap, but the separation and ordering of the segments is preserved.[3]

With respect to the second point, Ryle (1949) offers a way of conceptualizing the relation between the mental and the physical that avoids the problems consequent upon identifying the mental with covert processes taking place inside the head:

> When we describe people as exercising qualities of mind, we are not referring to occult episodes of which their overt acts and utterances are effects, we are referring to those overt acts and utterances themselves. (p. 25)
>
> When a person talks sense aloud, ties knots, feints or sculpts, the actions which we witness are themselves the things which he is intelligently doing...He is bodily active and mentally active, but he is not being synchronously active in two different "places," or with two different "engines." There is one activity, but it is susceptible of and requiring more than one kind of explanatory description. (pp. 50-51)

This way of characterizing intelligent action does not eliminate the requirement that linguistic utterances must be planned. Rather it eliminates the idea that covert processes are privileged in being mental or psychological, whereas overt actions are not. Instead, we may think of the talker's intended message as it is planned, uttered, specified acoustically, and perceived as being replicated intact across different physical media from the body of the talker to that of the listener.

An event theory of speech production must aim to characterize articulation of phonetic segments as overlapping sets of coordinated gestures, where each set of coordinated gestures conforms to a phonetic segment. By hypothesis, the organization of the vocal tract to produce a phonetic segment is invariant over variation in segmental and suprasegmental contexts. The segment may be realized somewhat differently in different contexts (for example, the relative contributions of the jaw and lips may vary over different bilabial closures [Sussman, MacNeilage, & Hanson, 1973]), because of competing demands on the articulators made by phonetic segments realized in an overlapping time frame. To the extent that a description of speech production along these lines can be worked out, the possibility remains that phonetic segments are literally uttered and therefore are available to be directly perceived if the acoustic signal is sufficiently informative. Research on a "task dynamic" model of speech production (e.g., Kelso et al., 1986, this volume; Saltzman, in press; Saltzman & Kelso, 1983) may provide at the very least an existence proof that systems capable of realizing overlapping phonetic segments nondestructively can be devised.

2.3 The Third Barrier: The Acoustic Signal Does Not Specify Phonetic Segments

Putting aside the question whether phonetic segments are realized nondestructively in articulation, there remains the problem that the acoustic signal does not seem to reflect the phonetic segmental structure of a linguistic communication. It need not, even if phonetic segments are uttered intact. Although gestures of the vocal tract cause disturbances in the air, it need not follow that the disturbances specify their causes. For many researchers, they do not. Figure 2 (from Fant & Lindblom [1961] and Cutting & Pisoni [1978]) displays the problem.

A spectrographic display of a speech utterance invites segmentation into "acoustic segments" (Fant, 1973). Visibly defined, these are relatively homogeneous intervals in the display. Segmentation lines are drawn where abrupt changes are noticeable. The difficulty with this segmentation is the relation it bears to the component phonetic segments of the linguistic utterance. In the display, the utterance is the name, "Santa Claus," which is composed of nine phonetic segments, but 18 acoustic segments. The relation of phonetic segments to acoustic segments is not simple as the bottom of Figure 2 reveals. Phonetic segments may be composed of any number of acoustic segments, from two to six in the figure, and most acoustic segments reflect properties of more than one phonetic segment.

How do listeners recover phonetic structure from such a signal? One thing is clear; the functional parsing of the acoustic signal for the perceiver is not one into acoustic segments. Does it follow that perceivers impose their own parsing on the signal? There must be a "no" answer to this question for an event theory devised from a direct-realist perspective to be viable. The perceived parsing must be in the signal; the special role of the perceptual system is not to create it, but only to select it.

The first point to be made in this regard is that there is more than one physical description of the acoustic speech signal. A spectrographic display suggests a parsing into acoustic segments, but other displays suggest other parsings of the signal. For example, Kewley-Port (1983) points out that in a spectrographic display the release burst of a syllable-initial stop consonant looks quite distinct from the formant transitions that follow it (for example, see the partitioning of /k/ in "Claus" in Figure 2). Indeed, research using the spectrographic display as a guide has manipulated burst and transition to study their relative salience as information for stop place (e.g., Dorman, Studdert-Kennedy, & Raphael, 1977). However, Kewley-Port's "running spectra" for stops (overlapping spectra from 20 ms windows taken at successive five ms intervals following stop release) reveal continuity between burst and transitions in changes in the location of spectral peaks from burst to transition.

It does not follow, then, from the mismatch between acoustic segment and phonetic segment, that there is a mismatch between the information in the acoustic signal and the phonetic segments in the talker's message. Possibly, in a manner as yet undiscovered by researchers but accessed by perceivers, the signal is transparent to phonetic segments.

If it is, two research strategies should provide converging evidence concerning the psychologically relevant description of the acoustic signal. The first seeks a description of the articulatory event itself--that is, of

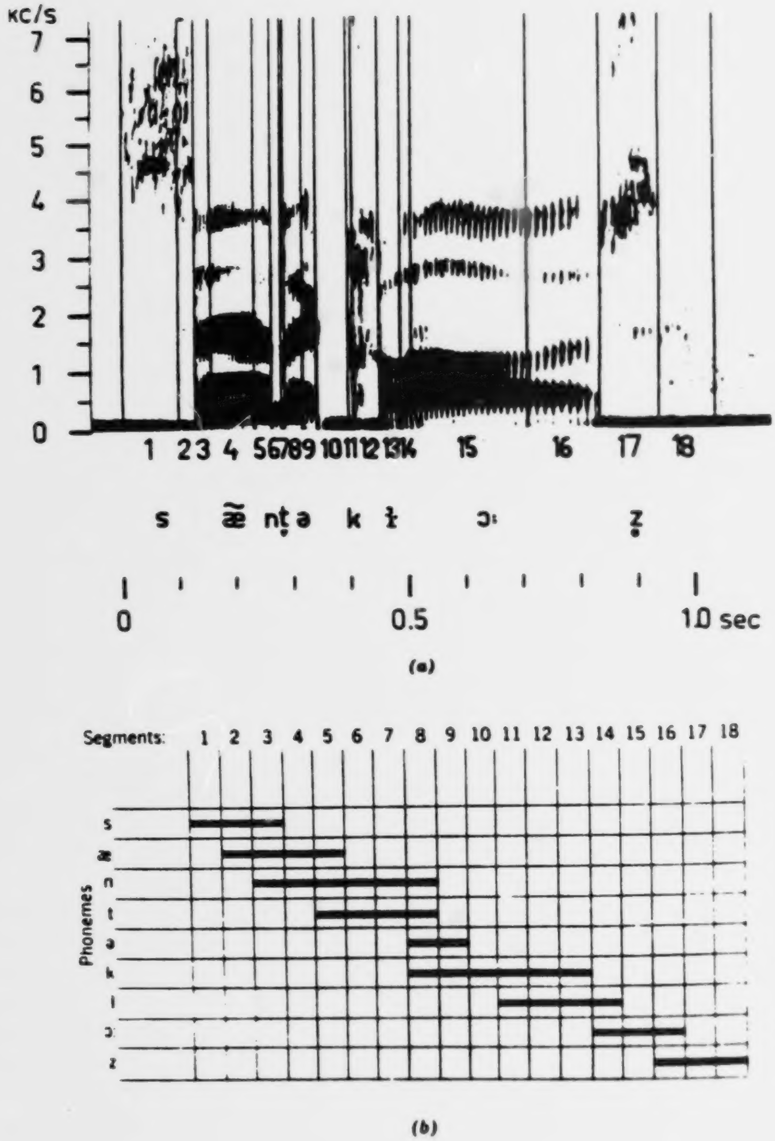


Figure 2: (a) Spectrographic display of "Santa Claus"; (b) Schematic display of the relationship between acoustic and phonetic segments (reprinted with permission from Cutting & Pisoni, 1978, and Fant & Lindblom, 1961).

sequences of phonetic segments as articulated--and then investigates the acoustic consequences of the essential articulatory components of phonetic segments. A second examines the parsing of the acoustic signal that listeners detect.

The research that comes closest to this characterization is that of Stevens and Blumstein (1978, 1981; Blumstein & Stevens, 1979, 1981). They begin with a characterization of phonetic segments and, based on the acoustic theory of speech production (Fant, 1960), develop hypotheses concerning invariant acoustic consequences of essential articulatory properties of the segments. They then test whether the consequences are, in fact, invariant over talkers and phonetic-segmental contexts. Finally, they ask whether these consequences are used by perceivers.

Unfortunately for the purposes of an event approach, perhaps, they begin with a characterization of phonetic segments as bundles of distinctive features. This characterization differs in significant ways from one that will be developed from a perspective on phonetic segments as coordinated articulatory gestures (see, for example, Browman and Goldstein, in press). One important difference is that the features tend to be static; accordingly, the acoustic consequences first sought in the research program were static also. A related difference is that the characterization deals with coarticulation by presuming that the listener gets around it by focusing his or her attention on the least coarticulated parts of the signal. As I will suggest shortly, that does not conform with the evidence; nor would it be desirable, because acoustic consequences of coarticulated speech are quite informative (cf. Elman & McClelland, 1983).

To date, Stevens and Blumstein have focused most of their attention on invariant information for consonantal place of articulation. Their hypotheses concerning possible invariants are based on predictions derived from the acoustic theory of speech production concerning acoustic correlates of constrictions in various parts of the vocal tract. As Stevens and Blumstein (1981) observe, when articulators adopt a configuration, the vocal tract forms cavities that have natural resonances, the formants. Formants create spectral peaks in an acoustic signal--that is, a range of frequencies higher in intensity than their neighbors. A constriction in the vocal tract affects the resonance frequencies and intensities of the formants. Thus, stop consonants with different places of articulation should have characteristic burst spectra independent of the vowel following the consonant and independent of the size of the vocal tract producing the constriction.

Blumstein and Stevens (1979) created "template" spectra for the stop consonants, /b/, /d/ and /g/, and then attempted to use them to classify the stops in 1800 CV and VC syllables in which the consonants were produced by different talkers in the context of various vowels. Overall, they were successful in classifying syllable-initial stops, but less successful with final stops, particularly if the stops were unreleased. Blumstein and Stevens (1980) also showed that listeners could classify stops by place better than chance when they were given only the first 10-46 ms of CV syllables.

However, two investigations have shown that the shape of the spectrum at stop release is not an important source of information for stop place. These studies (Blumstein, Isaacs, & Mertus, 1982; Walley & Carrell, 1983) pitted place information contributed by the shape of the spectrum at stop release in

CVs, against the (context-dependent) information for place contributed by the formant frequencies themselves. In both studies, the formants overrode the effect of spectral shape in listeners' judgments of place.

Recently, Lahiri, Gewirth, and Blumstein (1984) have found in any case that spectral shape does not properly classify labial, dental, and alveolar stops produced by speakers of three different languages. In search of new invariants and following the lead of Kewley-Port (1983), they examined the information in running spectra. They found that they could classify stops according to place by examining relative shifts in energy at high and low frequencies from burst to voicing onset. Importantly, pitting the appropriate running spectral patterns against formant frequencies for 10 CV syllables in a perceptual study, Lahiri et al. found that the spectral information was overriding. The investigators identify their proposed invariants as "dynamic," because they are revealed over time during stop release, and relational because they are based on relative changes in the distribution of energy at high and low frequencies in the vicinity of stop release.

Lahiri et al. are cautious whether their proposed invariants will withstand further test--and properly so, because the invariants are somewhat contrived in their precise specification. I suspect that major advances in the discovery of invariant acoustic information for phonetic segments will follow advances in understanding how phonetic segments are articulated. However, the proposals of Lahiri et al. (see also Kewley-Port, 1983) constitute an advance over the concept of spectral shape in beginning to characterize invariant acoustic information for gestures rather than for static configurations.

2.4 The Fourth Barrier: Perception Demonstrably Involves "Top Down" Processes and Perceivers Do Make Mistakes

Listeners may "restore" missing phonetic segments in words (Samuel, 1981; Warren, 1970), and talkers shadowing someone else's speech may "fluently restore" mispronounced words to their correct forms (e.g., Marslen-Wilson & Welsh, 1978). Even grosser departures of perceptual experience from stimulation may be observed in some mishearings (for example, "popping really slow" heard as "prodigal son" [Browman, 1980] or "mow his own lawn" heard as "blow his own horn" [Garnes & Bond, 1980]).

These kinds of findings are often described as evidence for an interaction of "bottom up" and "top down" processes in perception (e.g., Klatt, 1980). Bottom-up processes analyze stimulation as it comes in. Top-down processes draw inferences concerning stimulation based both on the fragmentary results of the ongoing bottom-up processes and on stored knowledge of likely inputs. Top-down processes can restore missing phonemes or correct erroneous ones in real words by comparing results of bottom-up processes against lexical entries. As for mishearings, Garnes and Bond (1980) argue that "active hypothesizing on the part of the listener concerning the intended message is certainly part of the speech perception process. No other explanation is possible for misperceptions which quite radically restructure the message..." (p. 238)

In my view (but not necessarily in the view of other event theorists), these data do offer a strong challenge to an event theory. It is not that an an event theory of speech perception has nothing to say about perceptual learning (for example, Gibson, 1966; Johnston & Pietrewicz, 1985). However, what is said is not yet well enough worked out to specify how, for example, lexical knowledge can be brought to bear on speech input from an direct-realist, event perspective.

With regard to mishearings, there is also a point of view (Shaw, Turvey, & Mace, 1982) that when reports of environmental events are in error, the reporter cannot be said to have perceived the events, because the word "perception" is reserved for just those occasions when acquisition of information from stimulation is direct and, therefore, successful. The disagreement with theories of perception as indirect and constructive, then, may reduce to a disagreement concerning how frequently bottom-up processes complete their work in the absence of top-down influence.

I prefer a similar approach to that of Shaw et al. that makes a distinction between what perceivers _can do_ and what they may do in particular settings. As Shaw et al. argue, there is a need for the informational support for activity to _be able to be_ directly extracted from an informational medium and for perception to be nothing other than direct extraction of information from proximal stimulation. However, in familiar environments, actors may generally guide their activities based not only on what they perceive, but also on what the environment routinely affords. In his presentation at the first event conference, Jenkins (1985) reviews evidence that the bat's guidance of flying sometimes takes this form. Placed in a room with barriers that must be negotiated to reach a food source, the bat soon learns the route (Griffin, 1958). After some time in which the room layout remains unchanged, a barrier is placed in the bat's usual flight path. Under these novel conditions, the bat is likely to collide with the barrier. Although it could have detected the barrier, it did not. By the same token, as a rule, we humans do not test a sidewalk to ensure that it will bear our weight before entrusting our weight to it. Nor do we walk through (apparent) apertures with our arms outstretched just in case the aperture does not really afford passage because someone has erected a difficult-to-see plate-glass barrier. In short, although the affordances that guide action _can be_ directly perceived, often they are not wholly. We perceive enough to narrow down the possible environments to one likely environment that affords our intended activity and other remotely likely ones that may not.

Perceptual restorations and mishearings imply the same perceptual pragmatism among perceivers of speech. It is also implied, I think, by talkers' tendencies to adjust the formality of their speaking style to their audience (e.g., Labov, 1972). Audiences with whom the talker shares substantial past experiences may require less information to get the message than listeners who share less. Knowing that, talkers conserve effort by providing less where possible.

It may be important to emphasize that the foregoing attempt to surmount the fourth barrier is intended to do more than translate a description of of top-down and bottom-up processes into a terminology more palatable to event theorists. In addition, I am attempting to allow a role for information not currently in stimulation to guide activity while preserving the ideas that perception itself must be direct and hence, errorless, and that activity _can be_ (but often is not) guided exclusively by perceived affordances.

As to the latter idea, the occurrence of mishearings that depart substantially from the spoken utterance should not deflect our attention from the observation that perceivers can hew the talker's articulatory line very closely if encouraged to do so. One example from my own research is provided by investigations of listeners' perceived segmentation of speech. Figure 1 above, already described, displays coarticulation of the primary articulators for vowels and consonants produced in a disyllable. This overlap has two general consequences in the acoustic signal (one generally acknowledged as a consequence, the other not). First, within a time frame generally identified with one phonetic segment (because the segment's acoustic consequences are dominant), the acoustic signal is affected by the segment's preceding and following neighbors. Second, because the articulatory trajectories for consonants overlap part of the trajectory of a neighboring vowel (cf. Carney & Moll, 1971; Öhman, 1966), the extent of time in the acoustic signal during which the vowel predominates in its effects--and hence the vowel's measured duration--decreases in the context of many consonants or of long consonants as compared to its extent in the context of few or short consonants (Fowler, 1983; Fowler & Tassinary, 1981; Lindblom & Rapp, 1973).

Listeners can exhibit sensitivity to the information for the overlapping phonetic segments that talkers produce in certain experimental tasks. In these tasks, the listeners use acoustic information for a vowel within a domain identified with a preceding consonant (for example, within a stop burst or within frication for a fricative consonant) as information for the vowel (Fowler, 1984; Whalen, 1984). Moreover, listeners do not integrate the overlapping information for vowel and consonant. Rather, they hear the consonant as if the vowel information had been factored from it (Fowler, 1985) and they hear the vowel as longer than its measured extent by an amount correlated with the extent to which a preceding consonant should have shortened it by overlapping its leading edge (Fowler & Tassinary, 1981).

These studies indicate that listeners can track the talker's vocal tract activities very closely and, more specifically, that they extract a segmentation of the signal into the overlapping phonetic segments that talkers produce, not into discrete approximations to phonetic segments and not into acoustic segments. Of course, this is as it must be among young perceivers if they are to learn to talk based on hearing the speech of others. But whether or not a skilled listener will track articulation this closely in any given circumstance may depend on the extent to which the listener estimates that he or she needs to in order to recover the talker's linguistic message.

3. Perception of Speech Events in an Expanded Time Frame: Sound Change

Two remarkable facts about the bottom tier of dually structured language are that its structure undergoes systematic change over time and that the sound inventories and phonetic processes of language reflect the articulatory dispositions of the vocal tract and perceptual dispositions of the ear (Lindblom et al., 1983; Locke, 1983; Ohala, 1981; Donegan & Stampe, 1979). There are many phonological processes special to individual languages that have analogues in articulatory-phonetic processes general to languages. For example, most languages have shorter vowels before voiceless than voiced stops (e.g., Chen, 1970). However, in addition, among languages with a phonological length distinction, in some (for example, German; see Comrie, 1980), synchronic or diachronic processes allow phonologically long vowels only before voiced consonants. Similarly, I have already described a general articulatory tendency for consonants to overlap vowels in production so that vowels are measured to shorten before clusters or long consonants more than before singleton consonants or short consonants. Compatibly, according to Elert (1964; cited in Lindblom & Rapp, 1973), in stressed syllables, Swedish short vowels appear only before long consonants or multiple consonants; long vowels appear before a short consonant or no consonant at all. In Yawelmani (see Kenstowiscz & Kisseberth, 1979), a long vowel is made short before a cluster. Stressed vowels also are measured to shorten in the context of following unstressed syllables in many languages (Fowler, 1981; Lehiste, 1972; Lindblom & Rapp, 1973; Nooteboom & Cohen, 1975). Compatibly, in Chimwi:ni (Kenstowiscz & Kisseberth, 1979), a long vowel may not, in general, occur before the antepenultimate syllable of a word.

These are just a few examples involving duration that I have gathered, but similar examples abound as do examples of other kinds. We can ask: how do linguistic-phonological processes that resemble articulatory dispositions enter language?

An interesting answer that Ohala (1981) offers to cover some cases is that they enter language via sound changes induced by systematic misperception by listeners. One example he provides is that of tonal development in "tone languages," including Chinese, Thai, and others. Tonal development on vowels may have been triggered by loss of a voicing distinction in preceding consonants. A consequence of consonant voicing is a rising tone on the following vowel (e.g., Hombert, 1979). Following a voiceless consonant, the tone is high and falling. Historical development of tones in Chinese may be explained as the listeners' systematic failure to ascribe the tone to consonant voicing--perhaps because the voicing distinction was weakening--and to hear it instead as an intentionally-produced characteristic of the vowel.

This explanation is intriguing because, in relation to the perspective on perceived segmentation just outlined, it implies that listeners may sometimes recover a segmentation of speech that is not identical to the one articulated by the talker. In particular, it suggests that listeners may not always recognize coarticulatory encroachments as such and may instead integrate the coarticulatory influences with a phonetic segment with which they overlap in time. This may be especially likely when information for the occurrence of the coarticulating neighbor (or its relevant properties as in the case of voicing in Chinese) is weakening. However, Ohala describes some examples where coarticulatory information has been misparsed despite maintenance of the conditioning segment itself. Failures to recover the talker's segmental parsing may lead to sound change when listeners themselves begin producing the phonological segment as they recovered it rather than as the talkers produced it.

Recent findings by Krakow, Beddor, Goldstein, and Fowler (1985) suggest that something like this may underlie an ongoing vowel shift in English. In English, the vowel /æ/ is raising in certain phonetic contexts (e.g., Labov, 1981). One context is before a nasal consonant. Indeed, for many speakers of English the /æ/ in "can," for example, is a noticeably higher vowel than that in "cad."

One hypothesis to explain the vowel shift in the context of nasal consonants is that listeners fail to parse the signal so that all of the influences of the nasalization on the vowel are identified with the coarticulatory influence of the nasal consonant. As Wright (1980) observes, the nasal formant in a nasalized vowel, is lower in frequency than F1 of /æ/. Integrated with F1 of /æ/ or mistakenly identified as F1, the nasal formant is characteristic of a higher vowel (with a lower F1) than F1 of /æ/ itself.

Krakow et al. examined this idea by synthesizing two kinds of continua using an articulatory synthesizer (Rubin, Baer, & Mermelstein, 1979). One continuum was a [bɛd] to [bæd] series (henceforth, the bed-bad series) created by gradually lowering and backing the height of the synthesizer's model tongue in seven steps. A second, [bɛnd] to [bænd], continuum (henceforth bend-band) was created in similar fashion, but with a lowered velum during the vowel and throughout part of the following alveolar occlusion. (In fact, several bend-band continua were synthesized with different degrees of velar lowering. I will report results on just one representative continuum.) Listeners identified the vowel in each series as spelled with "E" or "A." Figure 3a compares the responses to members of the bed-bad continuum with responses to a representative bend-band series. As expected, we found a tendency for subjects to report more "E"s in the bend-band series.

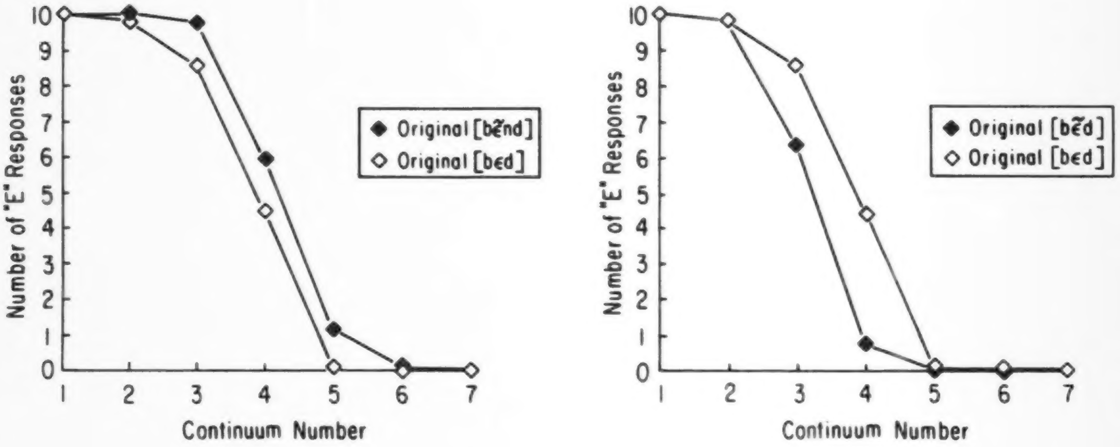


Figure 3: Identification of vowels in the experiment of Krakow, Beddor, Goldstein, and Fowler (1985); see text for explanation.

We reasoned that if this were due to a failure of listeners to parse the signal so that all of the acoustic consequences of nasality were ascribed to the nasal consonant, then by removing the nasal consonant itself, we would see as much or even more raising than in the context of a nasal consonant. Accordingly, we altered the original bed-bad series by lowering the model velum throughout the vowel. (I will call the new [bɛ̃d/-bæ̃d] continuum the bed(N)-bad(N) series. Again, different degrees of nasality were used over different continua. I will report data from a representative series.) Figure 3b shows the results of this manipulation. Rather than experiencing increased raising, as expected, the listeners experienced significant lowering of the vowel in the bed(N)-bad(N) series. Although this outcome can be rationalized in terms of spectral changes to the oral formants of the vowel due to the influence of the nasal resonance on them, it does not elucidate the origin of the raising observed in the first study.

A difference between our bend-band and bed-bad series was in the measured duration of the vowels. Following measurements of natural productions, we had synthesized syllables with shorter measured vowels in the bend-band series than in the bed-bad series. We next considered the possibility that this explained the raising we had found in the first experiment. /ɛ/ is an "inherently" shorter vowel than /æ/ (e.g., Peterson & Lehiste, 1960). It seemed possible that raising in the bend-band series was not due to misparsing of nasality, but to misparsing of the vowel's articulated extent from that of the overlapping nasal consonant. In particular, the vowels in the bend-band continua might have been perceived as inherently shorter (rather than as more extensively overlapped by the syllable coda) than vowels in the bed-bad series, and hence as more /ɛ/-like.

To test that idea, we synthesized a new bend-band series with longer measured durations of vowels, matching those in the original bed-bad (and bed(N)-bad(N)) series and new bed-bad and bed(N)-bad(N) series with vowels shortened to match the measured duration of those in the original bend-band series. Figures 4a and 4b show the outcome for the short and long series, respectively. Identification functions for bed-bad and bend-band are now identical. Listeners ascribe all of the nasality in the vowel to the consonant, and when vowels are matched in measured duration, there is no raising. Stimuli in the bed(N)-bad(N) series show lowering in both Figures 4a and 4b.

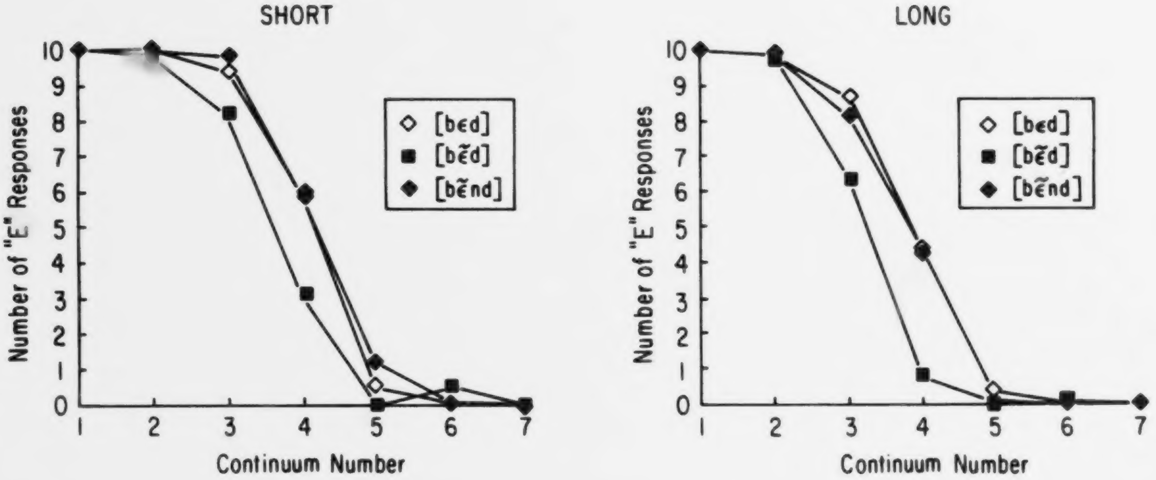


Figure 4: Identification of vowels in continua having vowels matched in measured duration (data from Krakow et al., 1985).

These results are of interest in several respects. For the present discussion, they are interesting in suggesting limitations in the extent to which these listeners could track articulation. Although listeners do parse speech along its coarticulatory lines in this study ascribing the nasality during the vowel to the nasal consonant, they are not infinitely sensitive to parts of a vowel overlaid by a consonant. The difficulty they have detecting the trailing edges of a vowel may be particularly severe when the following consonants are nasals as in the present example, because, during a nasal, the oral cavity is sealed off and the acoustic signal mainly reflects passage of air through the nasal cavity. Consequently, information for the vowel is poor. (There is vowel information in nasal consonants, however, as Fujimura, 1962, has shown.)

In a study mentioned earlier, Fowler and Tassinary found that in a vowel-duration continuum in which voicing of a final alveolar stop was cued by vowel duration (cf. Raphael, 1972), the "voiceless" percept was resisted more for vowels preceded by consonants that, in natural productions, shorten their measured extents substantially than by consonants that shorten them less. In the study, however, the effect on the voicing boundary was less than the shortening effect of the preceding consonant would predict. Together, this study and that by Krakow et al. suggest that although listeners do parse the speech signal along coarticulatory lines, they do not always hear the vowels as extending throughout their whole coarticulatory extent.[4]

As Ohala has suggested (1981), these perceptual failures may provoke sound change. Thereby they may promote introduction into the phonologies of languages, processes that resemble articulatory dispositions.

What are the implications of this way of characterizing perception and sound change for the theory of perception of speech events? In the account, perceivers clearly are extracting affordances from the acoustic signal. That is, they are extracting information relevant to the guidance of their own articulatory activities. (See the following section for some other affordances perceived by listeners.) However, just as clearly, the distal event they reported in our experiment and that they reproduce in natural settings is not the one in the environment. The problem here may or may not be the same as that discussed as the "fourth barrier" above. In the present case, the problem concerns the salience of the information provided to the listener in relation to the listener's own sensitivity to it. Information for vowels where consonants overlap them presumably is difficult (but not impossible, see Fowler, 1984; Whalen, 1984) to detect. One way to handle the outcome of the experiment by Krakow et al. within a direct-realist event theory is to suppose that listeners extract less information from the signal than they need to report their percept in an experiment or to reproduce it themselves, and they fill in the rest of the information from experience at the time of report or reproduction. An alternative is that listeners are insensitive to the vowel information in the nasal consonant (either because it is not there or because they fail to detect it) and use that lack of information as information for the vowel's absence there. Presumably it is just the cases where important articulatory information is difficult to detect that undergo the perceptually-driven sound changes in languages (cf. Lindblom, 1971).

4. How Perception Guides Action

Some Affordances of Phonetically-Structured Speech

For those of us engaged in research on phonetic perception, it is easy to lose sight of the fact that, outside of the laboratory, the object of perceiving is not the achievement of a percept, but rather the acquisition of information relevant to guidance of activity. I will next consider how perception of phonetically-structured vocal activity may guide the listener's behavior. This is not, of course, where most of the action is to be found in speech perception. More salient is that way the perception of the linguistic message guides the listener's behavior. This is a very rich topic, but not one that I can cover here.

Possibly, the most straightforward activity for a listener just having extracted information about how a talker controlled his or her articulators (but not, in general, the most appropriate activity), is to control one's own articulators in the same way--that is, to imitate. Indeed, research suggests that listeners can shadow speech with very short latencies (Chistovich, Klaas, & Kuzmin, 1962; Porter, 1977) and that their latencies are shorter to respond with the same syllable or one that shares gestures with it than with one that does not (Meyer & Gordon, 1984).

Although this has been interpreted as relevant to an evaluation of the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), it may also, or instead, reflect a more general disposition for listeners to mimic talkers (or perhaps to entrain to them). Research shows that individuals engaging in conversation move toward one another in speech rate (defined as number of syllables per unit time excluding pauses; Webb, 1972) in loudness (Black, 1949) and in average duration of pauses (Jaffe, 1964), although the temporal parameters of speaking also show substantial stability among individual talkers across a variety of conversational settings (Jaffe & Feldstein, 1974). In addition, Condon and Ogston (1971; also see Condon, 1976, for a review), report that listeners (including infants aged 1-4 days; Condon & Sander, 1974) move in synchrony with a talker's speech rhythms.

Although it is possible that this disposition for "interactional synchrony" (Condon, 1976) has a function, for example, in signaling understanding, empathy, or interest on the listener's part (cf. Matarazzo, 1965), the observations that some of the visible synchronies have been observed when the conversational partners cannot see one another, and some have been observed in infants, may suggest a more primitive origin. Condon (1976) suggests that interactional synchrony is a form of entrainment.

The disposition to imitate among adults may be a carryover from infancy, when presumably it does have an important function (Studdert-Kennedy, 1983). Infants must extract information about phonetically-structured articulation from the acoustic speech signals of mature talkers in order to learn to regulate their own articulators to produce speech. Although it seems essential that infants do this, very little research does more than hint that infants have the capacity to imitate vocal productions.

Infants do recognize the correspondence between visible articulation of others and an acoustic speech signal. They will look preferentially to the one of two videotapes on which a talker mouths a disyllable matching an accompanying acoustic signal (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). Moreover, infants recognize the equivalence of their own facial gestures to those of someone else. That is, they imitate facial gestures, such as lip or tongue protrusion (Meltzoff & Moore, 1985) even though, as Meltzoff and Moore point out, such imitation is "intermodal," because the infants cannot see their own gestures. Together, these findings suggest that infants should be capable of vocal imitation.

Relatively few studies have examined infants' imitation of adult vocalizations, however. Infants are responsive to mothers' vocalizations, and indeed, vocalize simultaneously with them to a greater-than-chance extent (Stern, Jaffe, Beebe, & Bennett, 1975). There are a few positive reports of vocal imitation (e.g., Kessen, Levine, & Wendrick, 1979; Kuhl & Meltzoff, 1982; Tuaycharoen, 1978; Uzgiris, 1973). However, few of them have been conducted with the controls now recognized as required to distinguish chance correspondences from true imitations.

Of course, imitative responses are not the only activities afforded by speech, even speech considered only as phonetically-structured activity of the vocal tract. A very exciting area of research in linguistics is on natural variation in speaking (e.g., Labov, 1966/1982, 1972, 1980). The research examines talkers in something close to the natural environments in which talking generally takes place. It is exciting because it reveals a remarkable sensitivity and responsiveness of language users to linguistically-, psychologically-, and socially-relevant aspects of conversational settings. Most of these aspects must be quite outside of the language users' awareness much of the time; yet they guide the talker's speech in quite subtle but observable ways.

Labov and his colleagues find that an individual's speaking style varies with the conversational setting in response, among other things, to characteristics of the conversational partner, including, presumably, the partner's own speaking style. Accordingly, adjustments to speaking style are afforded by the speech of the conversational partner.

An example of research done on dialectal affordances of the speech of other social groups is provided by Labov's early study of the dialects of Martha's Vineyard (1963). Martha's Vineyard is a small island off the coast of New England that is part of the state of Massachusetts. Whereas traditionally, residents were farmers and fishermen, in recent decades, the island has become a popular summer resort. The addition of some 40,000 summer residents to the year-round population of 5-6000 has, of course, had profound consequences for the island's economy.

Labov chose to study production of two diphthongs, [ɑi] and [ɑu], both of which had lowered historically from the forms [əi] and [əu]. These historical changes were not concurrent; [ɑu] had lowered well before the settlement of Martha's Vineyard by English speakers in 1642; [ɑi] lowered somewhat after its settlement.

Labov found a systematically increasing tendency to centralize the first vowel of the diphthongs--that is, to reverse the direction of sound change just described--in younger native residents when he compared speakers ranging in age from 30 upwards. The tendency to centralize the vowels was strongest among people such as farmers whose livelihoods had been most threatened by the summer residents. (The summer residents have driven up land prices as well as the costs of transporting supplies to the island and products to the mainland.) In addition, the tendency to centralize was correlated with the speaker's tendency to express resistance to the increasing encroachment of summer residents on the island. Among the youngest group studied, 15 year olds, the tendency to centralize the vowels depended strongly on the individual's future plans. Those intending to stay on the island showed a markedly stronger tendency to centralize the diphthongs than those intending to leave the island to make a living on the mainland. Labov interpreted these trends as a disposition among many native islanders to distinguish themselves as a group from the summer residents.

I find these data and others collected by Labov and his colleagues quite remarkable in the evidence they provide for listeners' responsiveness to phonetic variables they detect in conversation. In natural conversational settings, talkers use phonetic variation to psychological and social ends; and, necessarily given that, listeners are sensitive to those uses.

What Enables Phonetically-Structured Vocal-Tract Activity to Do Linguistic Work and How Is That Work Apprehended?

Confronted with language perception and use, an event theory faces powerful challenges. Gibson's theory of perception (1966, 1979) depends on a necessary relation between structure in informational media and properties of events. Obviously, physical law relating vocal tract activities to acoustic consequences satisfies that requirement well. But how is the relation between word and referent, and, therefore, between acoustic signal and referent to be handled? These relations are not universal--that is, different languages use different words to convey similar concepts. Accordingly, in one sense, they are not necessary and not, apparently, governed by physical law.

I have very little to offer concerning an event perspective on linguistic events (but see Verbrugge, 1985), and what I do have to say, I consider very tentative indeed. However, I would like to address two issues concerning the relation of speech to language. Stated as a question, the first issue is: what allows phonetically-structured vocal-tract activity to serve as a meaningful message? The second asks: can speech qua linguistic message be directly perceived?

As to the first question, Fodor (1974) observes that there are two types of answer that can be provided to questions of the form: "What makes X a Y?." He calls one type of answer the "causal story" and the other the "conceptual story." To use Fodor's example, in answer to the question: "What makes Wheaties the Breakfast of Champions?", one can invoke causal properties of the breakfast cereal, Wheaties, that turn nonchampions who eat Wheaties into champions. Alternatively, one can make the observation that disproportionate numbers of champions eat Wheaties. As Fodor points out, these explanations are distinct and not necessarily competing.

In reference to the question, what makes phonetically-structured vocal-tract activity phonological (that is, what makes it serve a linguistic function), one can refer to the private linguistic competences of speakers and hearers that allow them to control their vocal tracts so as to produce gestures having linguistic significance. Alternatively, one can refer to properties of the language user's "ecological niche" that support linguistic communication. Vocal-tract activity can only constitute a linguistic message in a setting in which, historically, appropriately constrained vocal-tract activity has done linguistic work. A listener's ability to extract a linguistic message from vocal-tract activity may be given a "conceptual" (I would say "functional") account along lines such as the following: Listeners apprehend the linguistic work that the phonetically-structured vocal tract activity is doing by virtue of their sensitivity to the historical and social context of constraint in which the activity is performed.

According to Fodor:

> Psychologists are typically in the business of supplying theories about the events that causally mediate the production of behavior...and cognitive psychologists are typically in the business of supplying theories about the events that causally mediate intelligent behavior. (p. 9)

He is correct; yet there is a functional story to be told, and I think that it is an account that event theorists will want to develop.

As to the second question, whether a linguistic message can be said to be perceived in a theory of perception from a direct-realist perspective, (direct) perception depends on a necessary relation between structure in informational media and its distal source. But as previously noted, this does not appear to apply to the relation between sign and significance.

Gibson suggests that linguistic communications, and symbols generally, are perceived (rather than being apprehended by cognitive processes), but indirectly. His use of the qualifier "indirect" requires careful attention:

> Now consider perception at second hand, or vicarious perception; perception mediated by communications and dependent on the "medium" of communication, like speech sound, painting, writing or sculpture. The perception is indirect since the information has been presented by the speaker, painter, writer or sculptor, and has been selected by him from the unlimited realm of available information. This kind of apprehension is complicated by the fact that direct perception of sounds or surfaces occurs along with the indirect perception. The sign is often noticed along with what is signified. Nevertheless, however complicated, the outcome is that one man can metaphorically see through the eyes of another. (1976/1982, p. 412).

By indirect, then, Gibson does not mean requiring cognitive mediation, but rather, perceiving information about events that have been packaged in a tiered fashion, where the upper tiers are structured by another perceiver/actor.

What is the difference for the perception of events that have a level of indirect as well as of direct specification? I do not see any fundamental difference in the manner in which perception occurs, although what is perceived is different. (That is, when I look at a table, it see it; when I hear a linguistic communication about a table, I perceive selected information about tables, not tables themselves.)

When an event is perceived directly, it is perceived by extraction of information for the event from informational media. When a linguistic communication is indirectly perceived, information for the talker's vocal-tract activities is extracted from an acoustic signal. The vocal-tract activity (by hypothesis) constitutes phonetically-structured words organized into grammatical sequences, and thereby indirectly specifies whatever the utterance is about.

It is worth emphasizing that the relation between an utterance (uttered in an appropriate setting) and what it signifies is necessary in an important sense. The necessity is not due to physical law directly, but to cultural constraints having evolved over generations of language use. These constraints are necessary in that anyone participating in the culture who communicates linguistically with members of the speech community must abide by them to provide information to listeners and must be sensitive to them to understand the speech of others.

Indeed, in view of this necessity, it seems possible that the distinction between direct and indirect perception could be dispensed with in this connection. Both the phonetically-structured vocal-tract activity and the linguistic information (i.e., the information that the talker is discussing tables, for example) are directly perceived (by hypothesis) by the extraction of invariant information from the acoustic signal, although the origin of the information is, in a sense, different. That for phonetic structure is provided by coordinated relations among articulators; that for the linguistic message is provided by constraints on those relations reflecting the cultural context of constraint mentioned earlier. What is "indirect" is apprehension of the table itself--which is not directly experienced; rather, the talker's perspective on it is perceived. Therefore, it seems, nothing is indirectly perceived.

I have attempted to minimize the differences between direct and "indirect" perception. However, there is a difference in the reliability with which information is conveyed. It seems that this must have to do with another sort of mediation involved in linguistic communications. As already noted, in linguistic communications the information is packaged into its grammatically structured form by a talker and not by a lawful relation between an event and an informational medium. And as noted much earlier, talkers make choices concerning what the listener already knows and what he or she needs to be told explicitly. Talkers may guess wrong. Alternatively, they may not know exactly what they are trying to say and therefore may not provide useful information. For their part, listeners, knowing that talkers are not entirely to be trusted to tell them what they need to know, may depend relatively more on extra-perceptual guesses.

References

Beattie, G. (1983). Talk: An analysis of speech and nonverbal behavior in conversation. Milton Keynes, England: Open University Press.

Black, J. W. (1949). *Loudness of speaking, I. The effect of heard stimuli on spoken responses.* Joint Project No 2 Contract N 7 Nmr-411 T. O. I., Project No NM 001 053 US Naval School of Aviation, Medicine and Research. Pensacola, Florida and Kenyon College, Gambier, OH (cited in Webb, 1972)

Blumstein, S. E., Isaacs, E., & Mertus, J. (1982). The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, *72*, 43-50.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production. *Journal of the Acoustical Society of America*, *66*, 1001-1017.

Blumstein, S. E., & Stevens, K. N. (1981). Phonetic features and acoustic invariance in speech. *Cognition*, *10*, 25-32.

Browman, C. P. (1980). Perceptual processing: Evidence from slips of the ear. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.

Browman, C. & Goldstein, L. (in press). Towards an articulatory phonology. *Phonology Yearbook*, Vol. 3, 1986.

Carney, P. J., & Moll, K. L. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, *23*, 193-202.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*, 129-159.

Chistovich, L., Klaas, I., & Kuzmin, I. (1962). The process of speech sound discrimination. Translated from *Voprosy Psikhologii*, *6*, 26-39.

Comrie, B. (1980). Phonology: A critical review. In B. Butterworth (Ed.), *Language production I*. London: Academic Press.

Condon, W. (1976). An analysis of behavioral organization. *Sign Language Studies*, *13*, 285-318.

Condon, W., & Ogston, W. (1971). Speech and body motion synchrony of the speaker-hearer. In D. Horton & J. Jenkins (Eds.), *The perception of language*. Columbus, OH: Charles C. Merrill.

Condon, W., & Sanders, W. (1974). Neonate movement is synchronous with adult speech: Interactional participation and language acquisition. *Science*, *183*, 99-101.

Cooper, A., Whalen, D. H., & Fowler, C. A. (1984). *Stress centers are not perceived categorically*. Paper presented to the 108th meeting of the Acoustical Society of America, Minneapolis, MN.

Donegan, P., & Stampe, D. (1979). The study of natural phonology. In D. Dinnsen (Ed.), *Current approaches to phonological theory*. Bloomington, IN: Indiana University Press.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, *22*, 109-122.

Elert, C. C. (1964). Ljud och ord i svenskan. Stockholm: Almqvist and Wiksell (Cited in Lindblom & Rapp, 1973).

Elman, J. and McClelland, J. (1983). Speech perception as a cognitive process: The interactive activation model. (ICS Report No 8302) San Diego: University of California, Institute of Cognitive Science.

Fant, G. (1960). *Acoustic theory of speech production*. 's Gravenhage: Mouton.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception & Psychophysics*, *27*, 343-350.

Fodor, J. A. (1975). *The language of thought*. New York: Thomas Y. Crowell.

Fowler, C. A. (1979). "Perceptual centers" in speech production and perception. *Perception & Psychophysics*, *25*, 375-388.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. Phonetica, 38, 35-50.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. Journal of Experimental Psychology: General, 112, 386-412.

Fowler, C. A. (1985). Segmentation of coarticulated speech in perception. Perception & Psychophysics, 36, 359-368.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes.Hillsdale, NJ: LEA.

Fowler, C. A. & Tassinary, L. Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), Attention and performance IX. Hillsdale, NJ: LEA.

Garnes, S., & Bond, Z. (1980). A slip of the ear: A snip of the ear? A slip of the year? In V. Fromkin (Ed.), Errors in linguistic performance: Slips of the tongue, ear, pen, and hand. New York: Academic Press.

Gibson, J. J. (1966). The problem of temporal order in stimulation and perception. Journal of Psychology, 62, 141-129. Reprinted in E. Reed & R. Jones (Eds.), Reasons for realism. Hillsdale, NJ: LEA, 1982.

Gibson, J. J. (1976). The theory of affordances and the design of the environment. (Unpublished). In E. Reed and R. Jones (Eds.), Reasons for realism. Hillsdale, NJ: LEA, 1982.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton-Mifflin.

Hammarberg, R. (1976). The metaphysics of coarticulation. Journal of Phonetics, 4, 353-363.

Hammarberg, R. (1982). On redefining coarticulation. Journal of Phonetics, 10, 123-137.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1, 1-7.

Hockett, C. (1985). Manual of phonology. (Publications in Anthropology and Linguistics, No. 11) Bloomington, IN: Indiana University Press.

Hockett, C. (1960). The origin of speech. Scientific American, 203, 89-96.

Hombert, J.-M. (1979). Consonant types, vowel quality and tone. In V. Fromkin (Ed.), Tone: A linguistic survey. New York: Academic Press.

von Hornbostel, E. M. (1927). The unity of the senses. Psyche, 7, 83-89.

Jaffe, J. (1964). Computer analyses of verbal behavior in psychiatric interviews. In D. Rioch (Ed.), Disorders of communication: Proceedings of the Association for Research in Nervous and Mental Diseases, Volume 42. Baltimore: Williams and Wilkins.

Jaffe, J., & Feldstein, S. (1970). Rhythms of dialogue. New York: Academic Press.

Javkin, H. (1976). The perceptual bases of vowel-duration differences associated with the voiced/voiceless distinction. Reports from the Phonetics Laboratory (Berkeley), 1, 78-89.

Jenkins, J. (1985). Acoustic information for objects, places and events. In W. Warren & R. Shaw (Eds.), Persistence and change. Hillsdale, NJ: LEA.

Johnston, T., & Pietrewicz, A. (Eds.). (1985). Issues in the ecological study of learning. Hillsdale, NJ: LEA.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory coopertion following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 812-832.

Kenstowiscz, M., & Kisseberth, C. (1979). *Generative phonology: Description and theory*. New York: Academic Press.

Kessen, W., Levine, J., & Wendrick, K. (1979). The imitation of pitch in infants. *Infant Behavior and Development*, *2*, 93-100.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*, 322-335.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: LEA.

Krakow, R. A., Beddor, P. S., Goldstein, L. M., & Fowler, C. A. (1985). *Effects of contextual and noncontextual nasalization on perceived vowel height*. Paper presented at the 109th Acoustical Society of America, Austin, TX.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138-1144.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*, 273-309.

Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics (third printing, 1982).

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania.

Labov, W. (Ed.). (1980). *Locating language in time and space*. New York: Academic Press.

Labov, W. (1981). Resolving the neogrammarian controversy. *Language*, 1981, 57, 267-308.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America*, *76*, 391-404.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, *51*, 2018-2024.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36.

Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University.

Lindblom, B. (1971). Phonetics and the description of language. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the 7th International Phonetics Congress* (pp. 63-97). The Hague: Mouton.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & D. Dahl (Eds.), *Universals workshop*. The Hague: Mouton.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics* (University of Stockholm), *21*, 1-59.

Lisker, L. (1978). *Rapid* vs. *Rabid*: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research*, *SR-54*, 127-132.

Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. Science, 219, 1347-1349.

MacNeilage, P. F., & Ladefoged, P. (1976). The production of speech and language. In E. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. 7): Language and perception. New York: Academic Press.

Marcus, S. (1981). Acoustic determinants of perceptual center (P-center) location. Perception & Psychophysics, 30, 247-256.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10, 29-63.

Matarazzo, J. D. (1965). The interview. In B. B. Wolman (Ed.), Handbook of clinical psychology. New York: McGraw-Hill.

Mattingly, I. G., Liberman, A. M., Syrdal, A. M., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. Cognitive Psychology, 2, 131-157.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

McNeill, D. (1985). So you think gestures are nonverbal? Psychological Review, 92, 350-371.

Meltzoff, A., & Moore, M. K. (1985). Cognitive foundations and social functions of imitation. In J. Mehler & R. Fox (Eds.), Neonate cognition. Hillsdale, NJ: LEA.

Meyer, D., & Gordon, P. (1984). Perceptual-motor processing of phonetic features in speech. Journal of Experimental Psychology: Human Perception and Performance, 10, 153-171.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). Psychological Review, 83, 405-408.

Neisser, U. (1967). Cognitive psychology. New York: Appleton-Century-Crofts.

Nooteboom, S. G., & Cohen, A. (1975). Anticipation in speech production and its implications for perception. In A. Cohen & S. G. Nooteboom (Eds.), Structure and process in speech perception. New York: Springer-Verlag.

Ohala, J. (1981). The listener as a source of sound change. In M. F. Miller (Ed.), Papers from the parasession on language behavior. Chicago: Chicago Linguistic Association.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 39, 151-168.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32, 693-703.

Porter, R. (1977). Speech production measures of speech perception: Systematic replications and extensions. Paper presented to the 93rd meeting of the Acoustical Society of America, Pennsylvania State University.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. Journal of the Acoustical Society of America, 51, 1296-1303.

Rapp, K. (1971). A study of syllable timing. Papers from the Institute of Linguistics (University of Stockholm), 14-19.

Repp, B. H. (1981). On levels of description in speech research. Journal of the Acoustical Society of America, 69, 1462-1464.

Rubin, P. E., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. Journal of the Acoustical Society of America, 70, 321-328.

Ryle, G. (1949). The concept of mind. New York: Barnes and Noble.

Saltzman, E. L. (in press). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Supplementum*.

Saltzman, E. L., & Kelso, J. A. S. (1983). Skilled actions: A task dynamic approach. *Haskins Laboratories Status Report on Speech Research*, *SR76*, 3-50.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474-494.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. MacNeilage (Ed.), *The production of speech*.New York: Springer-Verlag.

Shaw, R., & Bransford, J. (1977). Introduction: Psychological approaches to the problem of knowledge. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, NJ: LEA.

Shaw, R. E., Turvey, M. T., & Mace, W. (1983). Ecological psychology: The consequences of a commitment to realism. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes*.Hillsdale, NJ: LEA.

Stern, D., Jaffe, J., Beebe, B., & Bennett, S. (1975). Vocalizing in unison and in alternation: Two modes of communication within the mother-infant dyad. In D. Aaronson & R. Reiber (Eds.), *Developmental psychology and communication disorders Annals of the New York Academy of Sciences*, *263*, 89-100.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*, 1358-1368.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech* (pp. 1-38). Hillsdale, NJ: LEA.

Studdert-Kennedy, M. (1983). On learning to speak. *Human Neurobiology*, *2*, 191-195.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, *16*, 397-420.

Tuaycharoen, P. (1978). The babbling of a Thai baby: Echoes and responses to the sounds made by adults. In N. Waterson & C. Snow (Eds.), *The development of communication*. Chichester: John Wiley.

Tuller, B., & Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. *Perception & Psychophysics*, *27*, 277-283.

Uzgiris, I. (1973). Patterns of vocal and gestural imitation in infants. In L. J. Stone, H. T. Smith, & L. B. Murphy (Eds.), *The competent infant*. New York: Basic Books.

Verbrugge, R. R. (1985). Language and event perception: Steps toward a synthesis. In W. Warren & R. Shaw (Eds.), *Persistence and change*. Hillsdale, NJ: LEA.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*, 1011-1022.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*, 392-393.

Warren, W., & Shaw, R. (1985). Events and encounters as units of analysis for ecological psychology. In W. Warren & R. Shaw (Eds.), *Persistence and change*. Hillsdale, NJ: LEA.

Webb, J. (1972). Interview synchrony: An investigation of two speech rate measures. In A. W. Siegman & B. Pope (Eds.), Studies in dyadic communication. New York: Pergamon Press.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. Journal of the Acoustical Society of America, 69, 275-282.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. Perception & Psychophysics, 35, 49-64.

Wright, J. (1980). The behavior of nasalized vowels in the perceptual vowel space. Report from the Phonetics Laboratory (Berkeley), 5, 127-163.

Footnotes

[1] It may be useful to be explicit about the relationships among some concepts I will be referring to. Events are the primitive components of an "ecological" science--that is, of a study of actor/perceivers in contexts that preserve essential properties of their econiches. In the view of many theorists who engage in such studies (see, for example, the quotation from Shaw et al. above), the only viable version of a perceptual theory that can be developed within this domain is one that adopts a direct-realist perspective. I will take this as essential to the event (or ecological) approach, although, imaginably, a theory of the perception of natural events might be proposed from a different point of view.

[2] There are fundamental similarities between the view of speech perception from a direct-realist perspective and from the perspective of the motor theory. An important one is that both theories hold that the listener's percept corresponds to the talker's phonetic message, and that the message is best characterized in articulatory terms.

There are differences as well. As Liberman and Mattingly (1985) note, one salient difference is that the direct-realist theory holds that the acoustic signal is, in a sense, transparent to the perceived components of speech, while the motor theory does not. According to the motor theory, achievement of a phonetic percept requires special computations on the signal that take into account both the physiological-anatomical and the phonetic constraints on the activities of the articulators. A second difference is more subtle and perhaps will disappear as the theories evolve. Liberman and Mattingly propose that the objects of speech perception (at the level of description under consideration) are the "control structures" for observed articulatory gestures. Due to coarticulatory smearing, these control structures are not entirely redundant with the collection of gestures as they occur. My own view is that the smearing is only apparent, and, hence, the control structures are wholly redundant with the collections of articulatory gestures (properly described) constituting speech.

[3] This characterization may appear patently incorrect in cases where the same articulator is involved simultaneously in the production of more than one phonetic segment (for example, the tongue body during closure for [kh] in "key" and "coo" and the jaw during closure for [b] in "bee" and "boo"). However, Saltzman and Kelso (Saltzman, in press; Saltzman & Kelso, 1983) have begun to model this as overlapping but separate demands of different control structures on the same articulator and my own findings on perceived segmentation of speech (Fowler, 1984; see also Fowler & Smith, 1986) suggest that perceivers extract exactly that kind of parsing of the speech signal.

[4]Javkin (1976) has provided evidence for the opposite kind of error. In his research, listeners heard vowels as longer before voiced than voiceless consonants, perhaps because of the continuation of voicing during the consonant.

THE DYNAMICAL PERSPECTIVE ON SPEECH PRODUCTION: DATA AND THEORY*

J. A. S. Kelso,† E. L. Saltzman and B. Tuller†

Abstract. Presented here, in preliminary form, is a general theoretical framework that seeks to characterize the lawful regularities in articulatory pattern that occur when people speak. A fundamental construct of the framework is the *coordinative structure*, an ensemble of articulators that functions cooperatively as a single task-specific unit. Direct evidence for coordinative structures in speech is presented and a control scheme that realizes both the contextually-varying and invariant character of their operation is outlined. Importantly, the space-time behavior of a given articulatory gesture is viewed as the outcome of the system's dynamic parameterization, and the orchestration among gestures is captured in terms of intergestural phase information. Thus, both *time* and *timing* are deemed to be intrinsic consequences of the system's dynamical organization. The implications of this analysis for certain theoretical issues in coarticulation raised by Fowler (1980) receive a speculative, but empirically testable, treatment. Building on the existence of phase stabilities in speech and other biologically significant activities, we also offer an account of change in articulatory patterns that is based on the nonequilibrium phase transitions treated by the field of synergetics. Rate scaling studies in speech and bimanual activities are shown to be consistent with a synergetic interpretation and suggest a principled decomposition of languages. The CV syllable, for example, is observed to represent a stable articulatory configuration in space-time, thus rationalizing the presence of the CV as a phonological form in all languages (e.g., Clements & Keyser, 1983). The uniqueness of the present scheme is that stability and change of speech action patterns are seen as different manifestations of the same underlying dynamical principles--the phenomenon observed depends on which region of the parameter space the system occupies.

**Journal of Phonetics*, 1986, *14*, 29-59.

†Also Center for Complex Systems and Department of Psychology, Florida Atlantic University, Boca Raton.

Acknowledgment. This paper is dedicated to the daughter of the first and third authors, Kathleen Sasha Scott Kelso, born on July 7th, 1985. She cannot, however, be held responsible for any of the errors and speculations contained herein. We also thank Kevin Munhall for stimulating discussions on Section 4 and for performing the data analysis of the experiment described in Section 4.4. The authors' research is supported by NINCDS Program Project NS13617 and BRS Grant RR-05596, by U.S. Office of Naval Research (Psychological Sciences Division) Contract N00014-83-C-0083, and by a grant from The Stuttering Center, Baylor College of Medicine.

Though probably wrong, ambitious, and the outcome of much idle speculation, the simplicity of the present scheme is attractive and may offer certain unifying themes for the traditionally disparate disciplines of linguistics, phonetics, and speech motor control.

Prologue

The present paper represents, in part, a program of research that seeks to understand the lawful regularities that occur in articulatory patterns when people speak. The term dynamical in the title should not be interpreted as pure biomechanics. Rather, dynamics is used here in the fashion of Maxwell (1877), a forerunner of modern treatments of dynamical systems; namely, as the simplest and most abstract description of the forms of motion produced by a system. In a complex system like that of speech production, it is clearly impossible to investigate the behavior of each microscopic degree of freedom, however one defines them. The challenge of a dynamical approach is to identify and then lawfully relate macroscopic parameters (that operate on slow time scales) to the behavioral interactions among more microscopic articulatory components (that operate on faster time scales). A putative, but important advantage of a dynamical approach that, in principle, may allow for a unification of linguistics, phonetics, and speech motor control is the level-independent nature of dynamical description. Thus, dynamics can be specified at a global abstract level for a system of articulators as well as at the local, concrete level of muscle-joint behavior. The issue then becomes less one of translating a "timeless" symbolic representation into space-time articulatory behavior, as it is one of relating dynamics that operate on different intrinsic time scales. Obviously, this is only a way of posing the problem, but we believe--in the absence of evidence to persuade us otherwise--that it offers a principled solution.

1. Introduction

When a speaker produces a word, it is well known that the physical description of the word (whether acoustic, as displayed in a spectrogram or waveform, or articulatory, as in a cineradiographic sequence) varies widely with many factors. Among these are the rate at which the speaker talks, the word's pattern of syllabic stress or emphasis within an utterance, and the phonetic structure of surrounding words. The variations that arise as a consequence of such factors have long resisted unified systematic descriptions. Despite intensive research efforts, no one has sufficiently described either the acoustic or articulatory information that serves to specify a word in all its various contexts. Nor has anyone, to our knowledge, identified a canonical shape for a word and then transformed it, in a principled fashion, into the many other shapes that it may take.

Along with colleagues at Haskins Laboratories (e.g., Browman, Goldstein, Kelso, Rubin, & Saltzman, 1984) we are developing a solution to this problem by treating the units of language--conventionally described by linguists and phoneticians as, for example, phonemes, syllables, and words--as the product of time-invariant control structures for a system of vocal tract articulators. We assume that it is from the properties of these dynamically specified control structures, to be described presently, that the observed physical variations naturally arise.

A central hypothesis derived from the present theoretical framework is that articulators seldom move in an isolated, independent fashion (cf. Bernstein, 1967).[1] In speech production, they are coordinated with one another in such a way that changes over time in vocal tract shape are produced. Such changes in vocal tract shape structure the sounds of speech for a listener. A central problem for the theory, then, which we shall address in the present paper, becomes one of characterizing interarticulator cooperation in a multidegree of freedom system, and identifying the "significant informational units of action" (Greene, 1971, p. xviii) for speech. We and others have provided theoretical and empirical support for the hypothesis that, for skilled movements of the limbs or speech articulators, such action units (or coordinative structures) do not entail rigid or hardwired control of joint and/or muscle variables (e.g, Fowler, 1977; Kugler, Kelso, & Turvey, 1980; Turvey, 1977; for recent reviews see Kelso, in press; Kelso & Tuller, 1984a). Rather, they are defined more abstractly in a task-specific manner, and serve to marshall the articulators temporarily and flexibly into functional groupings or ensembles of joints and muscles that can accomplish particular goals. But what principles govern the assembly of coordinative structures for speech and how can such structures be explicitly modeled?

In Section 2 of the present paper we present evidence of coordinative structures in speech and discuss how they might be used in the production of single syllable utterances. Our focus is primarily on the task-specific stability of coordinative structures in the face of either experimentally-induced mechanical perturbations, or "natural" perturbations that might occur during ongoing speech as a result of contextual variations. A key feature of the model we are developing, termed task dynamics (e.g., Saltzman & Kelso, 1983/in press) is that it allows one to define invariant control structures for specific vocal tract gestural goals, from which contextually varying patterns of articulatory trajectories arise. These structures are invariant in two ways, both qualitatively, in terms of the set of relations among dynamic parameters (analogous to mass, stiffness, damping, etc.), and quantitatively, in terms of the parameter values themselves.

Speaking a word entails laryngeal and supralaryngeal gestures involving coordinated activity of many articulators. But words are not simply strings of individual gestures, produced one after the other; rather, each is a particular pattern of gestures, orchestrated appropriately in space and time. A way to probe the nature of the underlying ordering process is to induce naturally occurring scaling transformations, such as changes in speaking rate and degree of prosodic stress, and search for those aspects of the articulatory pattern that remain stable across these transformations. In Section 3 of the present paper we reconceptualize and perform an extensive reanalysis of earlier work that showed that the relative timing among articulators was a crucial feature of intergestural coordination. These steps point to the importance of phase as a key dependent variable, a finding that has empirical and theoretical implications for understanding both the stability of the spatiotemporal orchestration among gestures and the dynamical control structures that underlie such patterns (Kelso & Tuller, 1985a, 1985b). We outline a dynamical account of speech production[2] that differs radically from views that characterize speech as a planned sequence of static linguistic/symbolic units that are different in kind from the physical processes involved in the execution of such a plan. Rather, we hypothesize that the coordinative structures for speech are dynamically defined in a unitary way across both abstract "planning" and concrete articulatory "production" levels. These units are not timeless, but rather incorporate time in an intrinsic manner (cf. Bell-Berti & Harris, 1981; Fowler, 1980).

In the final section we discuss new directions--experimental and theoretical--for enlarging our understanding of the subtleties of dynamical structure that underlie changes in critically scaled articulatory patterns. We speculate that the form of such changes may in fact offer a window into, and perhaps even rationalize, the basic units of phonological analysis.

2. On Coordinative Structures in Speech

2.1 Theory and Data

The production of a single syllable requires the cooperation among a large number of neuromuscular components at respiratory, laryngeal, and supralaryngeal levels, operating on different time-scales. Yet somehow from this huge dimensionality the sound emerges as a distinctive and well-formed pattern. How this "compression" occurs--from a microscopic basis of huge dimensionality to a low-dimensional macroscopic description--is central to many realms of science, not only to understanding the coordination among speech articulators (see e.g., Kelso & Scholz, 1985). For example, there are many neurons, neuronal connections, metabolic components, muscles, motor units, etc. involved in pointing a finger at a target, yet the action itself is nicely modeled by a mass-spring system, a point attractor dynamic in which all system trajectories converge asymptotically at the desired target (e.g, Cooke, 1980; Kelso, 1977; Polit & Bizzi, 1978; Schmidt & McGown, 1980).

Is it, in fact, the case in speech that the higher dimensionality available actually reduces to a lower-dimensional, controllable system? If so, on what principles does such compression or reduction of the many degrees of freedom rest? These questions amount to a basic problem in the control of complex systems, that is, determining the circumstances under which a small set of control parameters (K) can effectively manipulate a much larger number of degrees of freedom (N). As Rosen (1980) notes, it is usually the case that K <<N, so that unless further constraints are imposed, it is not possible to impose arbitrary controls on N degrees of freedom.

But what form do such constraints take? Is there any evidence that the many degrees of freedom are actually constrained in a systematic fashion when a person talks? In earlier work, Fowler (1977, 1980) has described some mostly indirect evidence that the many neuromuscular components involved in speech do, in fact, cooperate to form functionally-specific action units, or as we prefer to call them, coordinative structures (e.g., Turvey, 1977). Here we supply more direct experimental support.

Support for the hypothesis that a group of relatively independent muscles and joints forms a single functional unit would be obtained if it were shown that a challenge or perturbation to one or more members of the group was, during the course of activity, responded to by other remote (non-mechanically linked) members of the group. We have recently found that speech articulators (lips, tongue, jaw) produce functionally specific, near-immediate compensation to unexpected perturbation, on the first occurrence, at sites remote from the locus of perturbation (Kelso, Tuller, & Fowler, 1982; Kelso, Tuller, V.-Bateson, & Fowler, 1984). The responses observed were specific to the actual speech act being performed: for example, when the jaw was suddenly perturbed while moving toward the final /b/ closure in /bæb/, the lips compensated so as to produce the /b/, but no compensation was seen in the tongue. Conversely, the same perturbation applied during the utterance /bæz/ evoked rapid and increased tongue muscle activity (appropriate for achieving a tongue-palate configuration for the final fricative sound) but no active lip compensation.

In order to explore the microscopic workings of a coordinative structure, recent work has also varied the phase of the jaw perturbation during bilabial consonant production. Remote reactions in the upper lip were observed only when the jaw was perturbed during the closing phase of the motion, that is, when the reactions were necessary to preserve the identity of the spoken utterance (see also Munhall & Kelso, 1985). Thus the form of cooperation observed is not rigid or "hard wired": the unitary process is flexibly assembled to perform specific functions (for additional evidence in speech and other activities, see Abbs, Gracco, & Cole, 1984; Berkenblit, Fel'dman, & Fukson, in press; Kelso et al., 1984). Elsewhere we have drawn parallels between these findings and brain function in general (Kelso & Tuller, 1984a). Just as groups of cells, not single cells, appear to be the main units of selection in higher brain function (Edelman & Mountcastle, 1978), so too task-specific ensembles of neuromuscular elements appear to be significant units of control and coordination of action, including speech.

To propose the coordinative structure as a fundamental unit of action does not just involve a change in terminology. Its purpose is to take us away from the hard-wired language of reflexes and central pattern generators (CPG) or the hard-algorithmed language of computers (formal machines), which is the source of the motor program/CPG idea. Reflexes and CPGs may be viewed as elemental entities, but they are not fundamental, in terms of affording an understanding of coherent action. The fact that we observe functionally-specific forms of cooperative behavior in many different creatures (e.g, the wiping behavior of the spinal frog; Berkenblit et al., in press; Fukson, Berkenblit, & Fel'dman, 1980) with vastly different neuroanatomies suggests that there may be nothing special, a priori, about neural structures and their "wiring" that mandates the existence of coordinative structures. Rather, it suggests that the functional cooperativity--not the neural mechanism per se--is fundamental. Although neural processes serve to instantiate such functions and support such cooperative behaviors, it is the lawful dynamical (rather than neural) basis of these cooperative phenomena that is our primary theoretical and experimental concern. This is where we part company with certain current views of motor control. Contrary to the motor programming formulation that relies on symbol-string manipulation familiar to computer technology (and, we would add, the whole "information processing" perspective of cognitive science [Carello, Turvey, Kugler, & Shaw, 1984]), the construct of coordinative structures highlights both the analytic tools of qualitative (nonlinear) dynamics (e.g., Kelso, Holt, Rubin, & Kugler, 1981; Kelso, V.-Bateson, Saltzman, & Kay, 1985; Saltzman & Kelso, 1983/in press), which provide low-dimensional descriptions of forms of motion produced by high dimensional systems, and the physical principles of cooperative phenomena (e.g., Haken, 1975; Haken, Kelso, & Bunz, 1985; Kelso, 1981; Kelso & Tuller, 1984a, 1984b; Kugler, et al., 1980; Kugler, Kelso, & Turvey, 1982), which account for the emergence of order and regularity in nonequilibrium, open systems. Though preliminary, both approaches will be apparent below and in following sections.

2.2 Task Dynamic Modeling

One way of trying to understand the operation of a coordinative structure is to model it. What type of model could generate, in a task-specific manner, the trajectories characteristic of normal unperturbed speech gestures and the spontaneous, compensatory behaviors discussed above? Here we discuss briefly how these issues of multiarticulator coordination within single speech gestures are treated in a task-dynamic model (Saltzman, 1985/in press; Saltzman & Kelso, 1983/in press) recently developed for effector systems having many articulatory degrees of freedom. Finally, we describe some preliminary attempts to model multiarticulator coordination within two temporally overlapping speech gestures, with reference to "naturally" induced compensatory behaviors (i.e., coarticulation).

Task dynamics is able to model the phenomenon of immediate compensation without requiring explicit trajectory planning or replanning (see Saltzman & Kelso, 1983/in press, for further details). Note that defining invariant patterns of dynamic parameters at the level of articulatory degrees of freedom (e.g., stiffness and damping parameters for the jaw and lips) will not suffice to generate these behaviors. The immediate compensation data for speech described above (Kelso et al., 1984) could not be generated by a system with a constant rest configuration parameter (i.e., a vector whose components are constant rest positions for the lips and jaw, cf. Lindblom, 1967). As shown in these data, when sustained perturbations were introduced during articulatory closing gestures, the system "automatically" achieved the same constriction as for an unperturbed gesture, but with a different final or rest configuration. Thus immediate compensation appears to result from the way that dynamic parameters at the articulatory level are constrained to change during a gesture in a context-dependent manner. In the task-dynamic model, such patterns of constraint originate in corresponding invariant patterns of dynamic parameters at an abstract, functionally defined level of task description.

There are three main steps involved in simulating coordinated movements of the speech articulators using the task-dynamic model. Since simulations to date have focused mainly on bilabial gestures, we will describe these three steps in some detail with reference to the specific example of a discrete bilabial closure task.

Task space. The first step is to specify the functional aspects of the given speech gesture with reference to the constriction-forming movements of an idealized vocal tract. This is done in a two-dimensional task space whose axes represent constriction location and constriction degree, and the topological form of the control regime for each task-space variable is specified according to the functional characteristics of the given speech task. For example, discrete and repetitive speech gestures will have damped (e.g., point attractor) and cyclic (e.g., limit cycle) second-order system dynamics, respectively, along each axis. At the task-space level, then, the dynamical system or control regime is abstract in that the constriction being controlled is independent of any particular set of articulators, and can refer, for example, to either a bilabial constriction produced by the lips and jaw or to a tongue-palate constriction produced by the tongue and jaw. Since we have chosen a discrete closure task to illustrate the steps involved in our task-dynamic simulations, we specify invariant, damped, second-order dynamics for the articulator-independent constriction along each task axis (see Figure 1a).

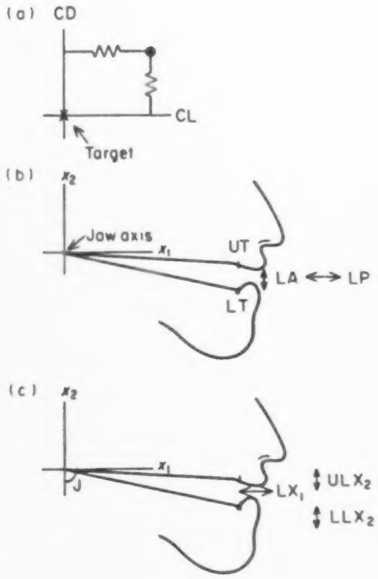


Figure 1. Bilabial tasks. a. Task space: variables are constriction location (CL) and constriction degree (CD). Closed circle denotes current system configuration. Sawtooth represents axis dynamics in lumped form. b. Body space: tract variables are lip protrusion (LP) and aperture (LA). UT and LT denote positions of upper and lower teeth, respectively. c. Model articulator space: articulator variables are jaw angle (J), upper lip vertical (ULX_2), lower lip vertical (LLX_2), and lip horizontal (LX_1).

Body space; *Tract variables*. The second step in modeling bilabial closure is to transform the task-space system kinematically into a two-dimensional body-space system defined in the midsagittal plane of the vocal tract. In contrast to the task-space regime, the body-space dynamics are specific to a given set of articulators whose movements govern the bilabial constriction along the *tract variable* dimensions of lip aperture (LA) and lip protrusion (LP). These tract variables represent the body-space counterparts of the task-space variables of constriction degree and location, respectively (see Figure 1b). Lip aperture is defined by the vertical distance between the upper and lower lips, and lip protrusion by the horizontal distances in the anterior-posterior direction of the upper and lower lips from the upper and lower teeth, respectively. Upper and lower lip protrusion movements are not independent in our preliminary formulation, but have been constrained to be equal in the model for purposes of simplicity. Consequently, like constriction location in task space, lip protrusion in body space currently constitutes only a single degree of freedom. This constraint may be abandoned in future work, as we attempt to model gestures in which the upper and lower lips show very different horizontal positions (e.g., labiodental fricatives).

Finally, it should be noted that the result of transforming from task space to body space coordinates is to define a two dimensional set of motion equations with a constant (although transformed) set of dynamic parameters. The tract variable control regimes are independent, since their corresponding equations of motion are uncoupled.

Model articulator space. The third step in modeling the closure task is to transform kinematically the two-dimensional tract variable regime into the coordinates of a four-dimensional model articulator space. The model articulators are moving segments that have lengths but are massless (see Figure 1c), and are defined with reference to the simplified articulatory degrees of freedom adopted in the Haskins Laboratories software articulatory speech synthesizer (Rubin, Baer, & Mermelstein, 1981). For bilabial gestures, the set of articulator movements associated with lip aperture includes rotation of the jaw and vertical displacements of the upper lip and lower lip relative to the upper and lower front teeth, respectively; for lip protrusion, the set of articulator movements includes (currently) yoked horizontal displacements in the anterior-posterior direction of the upper and lower lips relative to the upper and lower front teeth, respectively.

Since there are more model articulator variables than tract variables for the bilabial closure task, the model articulator system is *redundant* and the inverse kinematic transform from tract variables to model articulator coordinates is indeterminate (e.g., Saltzman, 1979). In order to deal with the indeterminacy or one-to-many property of this transformation, a weighted, least-squares optimality constraint is introduced in the form of a *weighted Jacobian pseudoinverse* transformation. This pseudoinverse has also been used in control schemes for robot arms that have a surplus number of degrees of freedom (i.e., the number of joints in the arm is greater than the number of task-relevant, spatial degrees of freedom for the hand, e.g., Benati, Gaglio, Morasso, Tagliasco, & Zaccaria, 1980; Klein & Huang, 1983; Whitney, 1972). Specifically, the pseudoinverse is a function of two matrix components--the *Jacobian* and *articulator weighting* matrices. The Jacobian matrix defines the transformation that relates motions of the articulators at their current configuration or posture to corresponding tract-variable motions of the bilabial constriction. The elements of the Jacobian matrix are nonlinear functions of the current articulatory posture. The elements of the articulator weighting matrix, however, are constant during a given gesture. In current modeling, a given set of articulator weightings constrains the motion of an articulator in direct proportion to the relative magnitude of the corresponding weighting element. Hence, different articulator weighting patterns are associated with different amounts of relative motion on the part of the four articulators responsible for controlling the tract variables of the bilabial constriction. In this sense, elements of the articulator weighting matrix used in the associated pseudoinverse define a further set of constant parameters for the bilabial constriction's equation of motion.

To summarize, in the task-dynamic model one may interpret the task-specific, coherent movements of the model articulatory system as resulting from the way that instantaneous tract-variable "forces" acting on a particular vocal tract constriction are distributed across the model articulators during the course of the tract variable gesture. At any given instant during this gesture, the partitioning is based on two factors:

a) the task-specific, constant set of articulator weightings and tract-variable dynamic parameters (e.g., lip aperture stiffness and damping); and

b) the current values of elements in the posturally dependent Jacobian matrix. Because these elements are functions of the current posture of the model articulators, the dynamic parameters defined at the level of the model articulator variables (e.g., stiffness and damping of the jaw, upper lip, and lower lip) are also functions of the evolving articulatory configuration.

Example 1: Discrete bilabial closures: Unperturbed gestures. Given a fixed set of dynamic parameter values for the tract variables of lip aperture and lip protrusion, and a set of initial positions and velocities for the jaw, upper lip, and lower lip, the equations of motion for the model articulators will generate a pattern of coordinated articulatory movements that will achieve the task goal (e.g., bilabial closure) specified for the tract variables. For an initial configuration corresponding to open and relatively unprotruded lips, and with initial articulator velocities of zero, these coordinated movements will reflect the evolving task-specific motions of the tract variables en route to their specified targets, with motion characteristics (e.g., speed, degree of overshoot, etc.) specified by the pattern of tract-variable dynamic parameter values. Assuming the system is not perturbed during its motion trajectory, the relative extents of movement for the jaw and lips will be specified by the relative values of the associated articulator weightings. Thus, one weighting pattern might correspond to predominant jaw motion, while a second weighting pattern might correspond to predominant vertical motion of the lips for a given lip aperture trajectory.

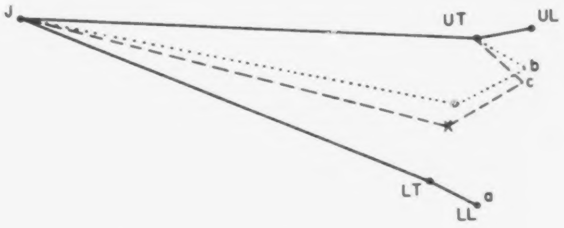


Figure 2. Simulated articulator configurations for bilabial closure task. a. Initial configuration (solid lines); b. Final configuration, unperturbed trajectory (dotted lines); c. Final configuration, perturbed trajectory (broken lines). Note that closure occurs lower in jaw space in c than in b. J = jaw axis, UT = upper teeth, UL = upper lip, LT = lower teeth, LL = lower lip.

Figure 2 (configurations a and b) illustrates an unperturbed movement from an initially open and relatively unprotruded configuration (Figure 2a) to a closed and relatively protruded final configuration (Figure 2b). Since the articulators associated with lip aperture were weighted equally in the corresponding weighting matrix, the extents of motion for these articulators were equal over the course of the gesture.

Example 2: Bilabial closure, immediate compensation, perturbed gestures. As discussed in Section 2.1, Kelso et al. (1984) demonstrated that if the jaw was retarded en route to a bilabial closure for /b/, the closure was still attained and the final articulatory configuration for the perturbed movement was different from the final configuration for unperturbed movements. Significantly, upper lip compensation was absent if the jaw was perturbed en route to an alveolar closure for /z/. These results show that an invariant dynamic description of a movement does not apply at the articulator level, since the articulatory-dynamic parameters (e.g., rest-configuration) must be able to change according to a movement's context in an utterance-specific (i.e., /b/ vs. /z/) manner. Furthermore, the speed of these compensatory behaviors suggests that they must occur "automatically" without reference to traditional stimulus-response reaction-time correction procedures.

The task-dynamic model handles such immediate compensation as follows. Bilabial closing gestures are simulated as discrete movements toward target constrictions, using point attractor dynamics for the local tract variables of lip aperture and protrusion. When the simulated jaw is "frozen" in place during the closing gesture at the level of the model effector system, the main qualitative features of the perturbation data are captured, in that: a) compensation to the jaw perturbation is immediate in the upper and lower lips, i.e., the system does not require reparameterization in order to compensate; and b) the target bilabial closure is reached (although with different final articulator configurations and, hence, different jaw-space locations for the closure) for both perturbed (Figure 2C) and unperturbed (Figure 2B) "trials."

Example 3: Coarticulation, gestural coproduction, bilabial and tongue dorsum gestures. In the task dynamic model, coarticulatory effects may originate in two ways. Passive carryover effects that are due to inherent system "sluggishness" (i.e., the time constants of the different tract variables; see also Henke, 1966; Coker, 1976) are implicit in the functioning of the model. Additionally, and more interestingly, other coarticulatory effects (both anticipatory and carryover) result from the temporally overlapping demands (conflicting or synergistic) made by the same or different tract variables on a common articulator subset (e.g., bilabial and tongue dorsum gestures with reference to the shared jaw articulator). We have begun to model these latter "active" coarticulatory effects by using the articulatory synthesizer to define articulator subsets for two new tract variables associated with the vocal tract constriction formed by movements of the tongue body dorsum. Thus, constriction location for tongue body dorsum is associated with the articulatory degrees of freedom of jaw rotation, and radial and angular displacements of the tongue body relative to the jaw; constriction degree for tongue body dorsum is associated with the same articulatory subset.

In preliminary simulations, we modeled the hypothetical cases in which bilabial and tongue dorsum gestures either did not overlap in time or were totally synchronous. Both gestures were identical in durational and damping factors, and all articulators had equal weightings. For both gesture types in the nonoverlapping case, the model articulators started at the same initial "neutral" configuration (corresponding to slightly open lips and a schwa-like position for the tongue dorsum), and attained their respective bilabial closure and tongue dorsum constriction targets. The final articulatory configurations were different for both gesture types and, in particular, the final jaw position for the single bilabial gesture was higher than that for the single tongue dorsum gesture. Recalling previous discussions in this section, these final configurational (and jaw positional) differences resulted from the different ways that the instantaneous, evolving task space forces were distributed across each gesture type's articulatory subset during the course of the movement. Roughly speaking, if we focus on the net "force" distributed to the jaw during the movement, we can say that more net force was delivered to the jaw during the simulated bilabial than during the tongue dorsum gesture, resulting in greater and lesser jaw displacements, respectively. Starting from the same initial configuration but with synchronous gestures, both the bilabial and tongue dorsum targets were again reached. However, the final articulatory configuration was different from those observed when either of the gestures occurred in isolation. The final jaw height for the gesturally synchronous case was halfway between the final jaw positions attained for the nonoverlapping gestures. This compromise jaw position resulted from the fact that, in the model, the net force delivered to the jaw over the gesturally synchronous movement was (roughly) the weighted average of the net jaw forces delivered during each of the nonoverlapping gestures.

We are extending our simulations currently to include cases in which different gestures overlap only partially in time (a more realistic assumption with reference to speech coarticulatory phenomena). In these cases, the net force distributed to the jaw (and hence total jaw displacement) during periods of gestural overlap will reflect the weighted averages of the jaw forces associated with each gesture over these periods. The predicted behavior of the model is consistent, in fact, with coarticulation data for V1CV2 utterances presented by Sussman, MacNeilage, and Hanson (1973). As a first approximation toward modeling such utterances, we will treat bilabial consonants as closing gestures of a lip-jaw system associated with the tract variables for bilabial constrictions. Similarly, we will treat vowels as opening gestures of the jaw-tongue system associated with the tract variables for tongue dorsum constrictions. We realize, of course, that this description represents only a preliminary, simplified account of the data, which will be modified as experiments and simulations progress. For example, at least the early portions of consonantal release gestures appear to depend on the manner class (e.g., stops vs. fricatives) of the consonants themselves. However, given these assumptions, we may represent the V1CV2 productions as temporally overlapping sequences of opening (vocalic) and closing (consonantal) tract variable gestures. Since the vowel and consonant gestures share the jaw as a common articulator, the net movement of the jaw during periods of gestural overlap (i.e., the period of jaw motion during which the V1C closing gesture overlaps the CV2 opening gesture) will be determined by the weighted average of the respective "demands" made on the jaw by each gesture during these periods. Hence, for example, the vertical upward displacement of the jaw for a V1C gesture (and hence, the jaw height at closure) will be influenced by the height of V2. Specifically, the net upward demand or "force" delivered to the jaw for low V2 (/æ/) will be less during the period of gestural overlap than it would be for high V2 (/i/), and should generate the anticipatory coarticulatory effect of greater V1C displacement for high V2 than for low V2 observed by Sussman et al. (1973).

3. On Gestural Orchestration: From Relative timing to Phase Stability.

In the previous section we focused on the intrinsic properties of functional units of action, but have not discussed the sequencing or orchestration of these units over time. One way to explore the processes underlying such orchestration is to transform a given action pattern as a whole (e.g., by scaling on movement rate, amplitude, etc.) and search for what remains stable across the transformation.

Much evidence now exists that the relative timing of movement events is stable across certain scaling changes and hence provides a more appropriate metric than their absolute durations. Although early demonstrations of relative temporal stability were provided from activities that are qualitatively repetitive and potentially pre-wired (e.g., locomotion, respiration, and mastication; see Grillner, 1977, for review), more recent work has revealed that less repetitive activities show similar organizational features (e.g., two-handed movements, typing, handwriting, postural control, and speech-manual coordination; Hollerbach, 1981; Kelso, Southard, & Goodman, 1979; Kelso, Tuller, & Harris, 1983; Lestienne, 1979; Nashner, 1977; Schmidt, 1982; Shapiro, Zernicke, Gregor, & Diestel, 1981; Viviani & Terzuolo, 1980). Importantly, there is some limited evidence that the production of speech can be described by a similar style of organization, and we will now describe this work in some detail.

In a set of previous experiments (Harris, Tuller, & Kelso, 1986; Tuller & Kelso, 1984; Tuller, Kelso, & Harris, 1982, 1983), Tuller and colleagues have shown that, across variations in speaking rate and stress, the timing of articulatory events associated with consonant production remains stable relative to the interval between events associated with flanking vowels. Consider a very simple, but paradigmatic case in which the latency (in ms) of onset of upper lip motion for a medial consonant is measured relative to the interval (in ms) between onsets of jaw motion for flanking vowels.

In Figure 3(top), we see the particular intervals measured for one token of the utterance /baPAB/, spoken at a conversational rate with primary stress on the second syllable. The movement data were obtained by recording from infrared LEDs attached to the subject's lips and jaw. Here, the interval from V1 to V2 represents the time between onsets of jaw motion for successive vowels. The interval V1-UL represents the latency of onset of medial consonant-related movement in the upper lip. These points were obtained from zero crossings of velocity traces. The main empirical question was: Do the intervals V1-V2 and V1-UL change in a systematically related way as syllable stress and speaking rate vary?

Figure 3(bottom), taken from Tuller and Kelso (1984), plots the latency of upper lip movement relative to the vowel period for one of the four speakers. The data were similar for all subjects. The utterance shown, /bapab/, spoken at two rates and with two stress patterns, illustrates the main result: over changes in speaking rate and stress, the measured temporal intervals and articulatory displacements change considerably, but the relative timing is preserved. The overall relationship can be described by a linear function defined by two parameters--a positive slope and a nonzero intercept. This high correlation of two event durations across rate and stress in different speakers has since been replicated by other investigators (Bladon & Al-Bamerini, personal communication; Gentil, Harris, Horiguchi, & Honda, 1984; Linville, 1982; Lubker, 1983; Munhall, in press).

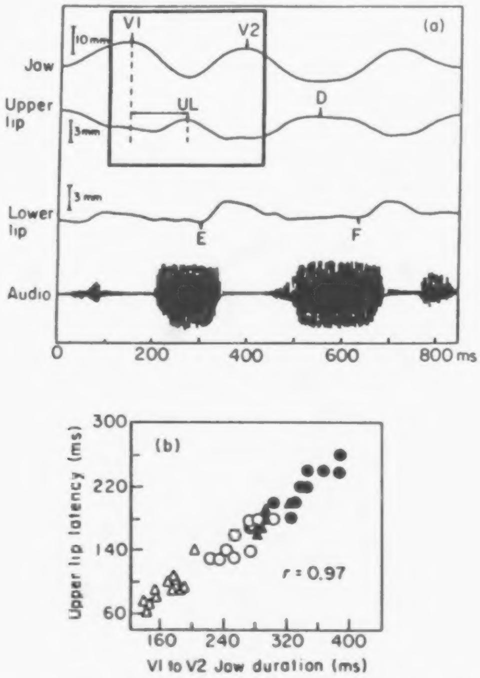


Figure 3. Top. Movements of the jaw, upper lip, and lower lip corrected for jaw movement, and the acoustic signal, for one token of /ba#'pab/. Articulator position (y-axis) is shown as a function of time. Onsets of jaw and lip movements are indicated (empirically determined from zero crossings in the velocity records). Bottom. Timing of upper lip lowering associated with /p/ production as a function of the period between successive jaw lowerings for the flanking vowels for one subject's productions of /ba#pab/. (●) Slow rate, first syllable stressed; (O) Slow rate, second syllable stressed; (▲) Fast rate, first syllable stressed; (Δ) Fast rate, second syllable stressed.

How is this stability of relative timing to be rationalized? A popular view in the motor control literature is that time is metered out by a central program that instructs or commands the articulators when to move, how far to move, and for how long (e.g., Schmidt, 1982). However, a reconceptualization by Kelso and Tuller (1985a/in press) and subsequent reanalysis of the original data (Tuller & Kelso, 1984) strongly suggest that their findings can be understood without recourse to an extrinsic timer or timing metric.[3] In fact, a very different view of articulatory "timing" emerges when the articulatory movements are reanalyzed as trajectories on the phase plane. These phase plane trajectories provide a geometric or kinematic description that usefully captures the forms of patterned motion produced by the articulators. A brief tutorial follows.

3. *The Phase Portrait: A Tutorial* (cf. Kelso, Tuller, & Harris, 1984a/1986)

All possible system states can be represented in the phase plane, whose axes are the articulator's position (x) and its velocity ($\dot{x}$). As time varies, the point P (x, $\dot{x}$) describing the motion of the articulator moves along a certain path on the phase plane. Figure 4 illustrates the mapping from time domain to phase plane trajectories. Hypothetical jaw and upper lip trajectories (position as a function of time) are shown for an unstressed /bab/ (Figure 4a, left) and a stressed /bab/ (Figure 4b, left). On the right are shown the corresponding phase plane trajectories. In this figure and those following we have reversed the typical orientation of the phase plane so that position is shown on the vertical axis and velocity on the horizontal axis. Thus, downward movements of the jaw are displayed as downward movements of the phase path. The vertical crosshair indicates zero velocity and the horizontal crosshair indicates zero position (midway between minimum and maximum displacement). As the jaw moves from its highest to its lowest point (from A to C in Figure 4), velocity increases (negatively) to a local maximum (B), then decreases to zero when the jaw changes direction of movement (C). Similarly, as the jaw is raised from the low vowel /a/ into the following consonant constriction, velocity peaks approximately midway through the gesture (D), then returns to zero (A).

Phase plane trajectories preserve some important differences between stressed and unstressed syllables. For example, maximum lowering of the jaw for the stressed vowel is greater than lowering for the unstressed vowel and maximum articulator velocity differs noticeably between these two orbits (e.g., Kelso et al., 1985; MacNeilage, Hanson, & Krones, 1970; Stone, 1981; Tuller, Harris, & Kelso, 1982). In contrast, the different *durations* taken to traverse the orbit as a function of stress are not represented explicitly in this description. That is, although time is implicit and usually recoverable from phase plane trajectories, it does not appear explicitly.

It is possible to transform the Cartesian $x,\dot{x}$ coordinates into equivalent polar coordinates, namely, a phase angle, $\phi = \tan^{-1}[\dot{x}/x]$, and a radial amplitude, $R = [x^2 + \dot{x}^2]^{1/2}$. These polar coordinates are indicated on the phase planes shown in Figure 4. The phase angle has been a key (computed) dependent variable in our re-analysis of interarticulator timing.[4] It allows us to rephrase the traditional question of how the lip knows *when* to begin its movement for the medial consonant by asking *where* on the cycle of jaw states that the lip motion for medial consonant production begins. One possibility is that lip motion begins at the same phase angle of the jaw across different jaw motion orbits (i.e., across rate and stress). This outcome is not necessarily entailed, or predicted by, the relative timing results. For example, Figure 4a through 4c shows three utterances whose vowel-to-vowel periods and consonant latencies do not change in a linearly related fashion. Nevertheless, the phase angle at which upper lip motion begins relative to the cycle of jaw states is identical in the three cases. Thus, the information for "timing" of a remote articulator (e.g., the upper lip) may not be time itself, nor absolute position of another articulator (e.g., the jaw), but rather a relationship defined over the position-velocity *state* (or, in polar coordinates, the phase angle) of the other articulator. Although this conceptualization is intriguing, we want to re-emphasize that it constitutes an *alternative description* of the relative timing data set. For example, Figure 5 illustrates the converse of Figure 4, namely, that two (hypothetical) utterances with identical vowel-to-vowel periods (P) and consonant latencies (L) can nonetheless show very different phase angles for upper lip movement onset. To be specific, the phase angle analysis incorporates the full trajectory of motion; the relative timing analysis is independent of trajectory once movement has begun and is based only on the onsets and offsets of movement events.

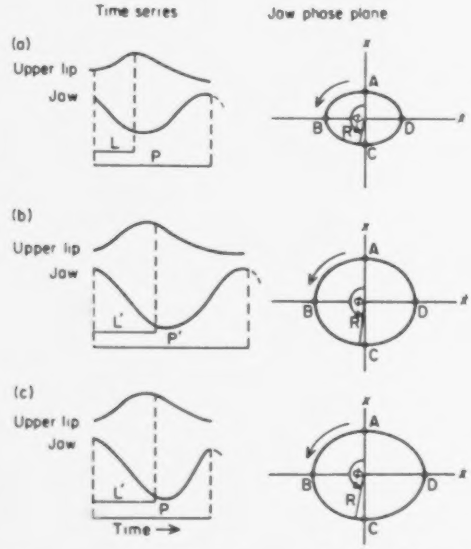

Time series
Jaw phase plane
(a)
Upper lip
Jaw
L
P
A
B
R
C
D
(b)
Upper lip
Jaw
L'
P'
A
B
R
C
D
(c)
Upper lip
Jaw
L'
P
A
B
R
C
D
Time →

Figure 4. Left: Time series representations of idealized utterances. Right: Corresponding jaw motions, characterized as a simple mass spring and displayed on the 'functional' phase plane (i.e., position on the vertical axis and velocity on the horizontal axis). Parts a, b, and c, represent three tokens with vowel-to-vowel periods (P and P') and consonant latencies (L and L') that are not linearly related. Phase position of upper lip movement onset relative to the jaw cycle is indicated (see text). (From Kelso & Tuller, 1985a/in press).

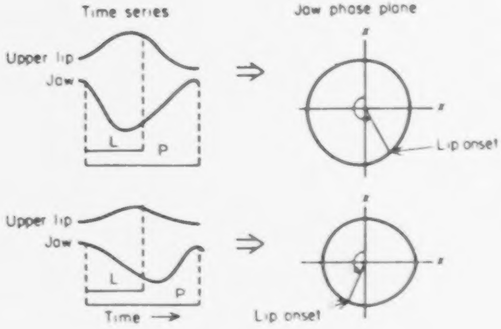

Time series
Jaw phase plane
Upper lip
Jaw
L
P
⇒
Lip onset
Upper lip
Jaw
L
P
Time →
⇒
Lip onset

Figure 5. Two hypothetical utterances having identical vowel-to-vowel periods (P) and consonant (upper lip) latencies (L) but different phase angles of upper lip onset. See caption Figure 4. (From Kelso & Tuller, 1985a/in press)

Figure 6 shows jaw motion on the phase plane for the first cycle of /'ba#bab/ (top) and /ba#'bab/ (bottom) produced at a fast rate. Each token shown is the first instance produced of the utterance type. On the left is the entire jaw cycle for each stress pattern; on the right, the jaw cycle is reproduced only until the point of onset of upper lip movement downward for production of the medial bilabial consonant, as measured from the first deviation from zero velocity. The calculated phase angle[5] at which upper lip motion begins is indicated for each token. Notice that the jaw displacement and velocity are both greater for the stressed than the unstressed syllable. Nevertheless, upper lip motion begins at essentially the same phase angle for both tokens. If upper lip motion began at a phase angle of 180°, it would be synchronous with the jaw "turnaround" point.

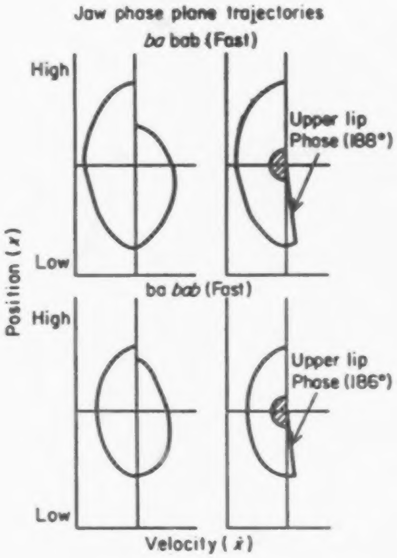


Figure 6. Left: Jaw cycle on the phase plane for the first token produced of stressed /ba#b/ (top) and unstressed /b#ab/ (bottom), spoken at a fast rate. Right: Jaw cycle until the onset of upper lip lowering for the second /b/. (From Kelso & Tuller, 1985a/in press)

3.2 New Results

Table 1 shows the mean data and the standard error of the mean for all four speakers from the Tuller and Kelso (1984) study. 2 X 2 ANOVAs for each utterance type showed no significant main effects of rate or stress or their interaction on the phase angle of upper lip onset for medial consonant production. For /babab/, *F*s (1,27) ranged from .02 to 2.97; for /bapab/, *F*s (1,30) ranged from .01 to 2.39; for /bawab/, *F*s (1,29) ranged from .01 to 2.80, *p*s > .1. Although phase angle was invariant across speaking rate and stress, Table 1 also shows some differences in phase angle as a function of the medial consonant. There is some tendency for upper lip phase for /p/ to be smaller than /b/. This result may be consistent with acoustic findings

Table 1

Mean upper lip phase (±SE) relative to vowel-to-vowel jaw trajectory for subjects JE, NM, BT, and CH

JE	/baba/	/bapa/	/bawa/
SS*	212 (7.05)	184 (6.41)	207 (3.69)
SU	205 (2.83)	183 (2.07)	205 (4.43)
FS	197 (2.83)	177 (3.82)	212 (6.08)
FU	203 (4.26)	179 (3.81)	203 (6.34)
NM			
SS	182 (2.0)	178 (2.35)	193 (2.19)
SU	178 (2.74)	175 (3.21)	193 (2.19)
FS	184 (3.14)	176 (1.73)	197 (2.30)
FU	183 (3.49)	172 (2.50)	189 (3.93)
BT			
SS	168 (2.76)	163 (2.93)	196 (2.91)
SU	168 (5.83)	174 (3.50)	198 (5.65)
FS	166 (4.58)	166 (2.68)	192 (5.23)
FU	164 (3.36)	167 (3.96)	191 (4.11)
CH			
SS	184 (4.38)	188 (6.45)	203 (5.48)
SU	186 (3.93)	184 (3.01)	208 (3.67)
FS	181 (6.38)	183 (3.94)	207 (5.90)
FU	182 (3.16)	177 (4.08)	196 (4.37)

*SS = Slow (Normal) speaking rate, first syllable stressed
SU = Slow (Normal) speaking rate, first syllable unstressed
FS = Fast Speaking Rate, first syllable stressed
FU = Fast Speaking Rate, first syllable unstressed

that vowels are longer before voiced than voiceless consonants. There is also a strong tendency for /w/'s upper lip phase to be greater than the stops. However, this result could be artifactual: the movement measures did not include a horizontal component (potentially larger for /w/ than /b/ or /p/). In addition, our subjects produced /w/ with much smaller and slower upper lip movements, making measurement of movement onset more difficult.

3.3 Empirical and Theoretical Implications

There are at least two empirical advantages of these phase angle analyses over our previous relative timing descriptions. First, in the relative timing analysis, the overall correlations across rate and stress conditions were very high, but the within-condition slopes tended to vary somewhat. In the phase analysis, on the other hand, the mean phase angle is the same across conditions. Second, recall that the relative timing data were fitted by linear functions described by two parameters. The phase description requires only a single parameter and, if nothing else, is the more parsimonious description.

The phase angle conceptualization also offers a number of theoretical advantages over the original relative timing analyses. First, once articulatory motions are represented geometrically on the phase plane, the phase angle serves to normalize duration across speaker, stress, speaking rate, etc. Second, these analyses potentially provide a grounding for so-called intrinsic timing theories of speech production (e.g., Fowler, 1980; Fowler, Rubin, Remez, & Turvey, 1980), since neither absolute nor relative durations need be monitored or controlled extrinsically, and no time-keeping mechanisms or time controllers are required in this formulation. As with a candle (which provides a metric for time by a change in its length) or a water clock (where the metric is number of drops), the units of time are defined entirely in terms of the dynamical processes involved. Time itself is not a fundamental variable, and is not likely to be a possessed, programmed, or represented property of the speech production system (Kelso, Tuller, & Harris, 1984; Kelso & Tuller, 1985a). As an aside, it has never been clear how the speech system could keep track of time, at least peripherally, because there is no known afferent basis (such as time receptors) for time-keeping in the articulatory structures themselves (Kelso, 1978). On the other hand, an informational basis (e.g., in position and velocity sensitivities of muscle spindle and joint structures) is a physiological given in the phase angle characterization. It might well be the case that certain critical phase angles provide information for coordination between articulators (beyond those considered here) and/or vocal tract configurations, just as phase angles of the leg joints provide coupling information for locomotory coordination (Shik & Orlovskii, 1965). Third, as Fowler (1980) notes, a dominant assumption of what she calls extrinsic timing theories of coarticulation is that phonological segments are considered to be discrete in the sense that their boundaries are straight lines perpendicular to the time axis. Yet as is well known, discrete segments are not seen by perpendicular cuts of the physical records of speech (acoustic, kinematic, physiological measurements) along the time axis. In the phase angle analysis, however, no *a priori* assumptions are made regarding the issue of segmentation *per se*, and the overlap (or coproduction) among gestures is captured in a natural way while still preserving a separation between consonantal and vocalic events.

A final implication of the view presented here is that "segments" or phonological units as typically defined by linguists may not be relevant to the speech production system. Rather, phonological units might be profitably reconceptualized in terms of characteristic interarticulator phase structures (see also Browman & Goldstein, in press, for related notions). Note that the phase structure description minimizes the mind/body problem for speech production by avoiding the translation step between psychological planning units and the physical execution of those units. On the other hand, different issues are immediately raised--such as whether there are a restricted number of stable phase structures (which one might expect if they are to be tagged with linguistic descriptors), and if so, why some configurations appear and not others. Experimental inroads into these issues can be made in the present perspective with a minimum of *ad hoc* assumptions, and with little resort to *a priori* linguistic categories.

4. Instabilities; Nonequilibrium Phase Transitions and Phonetic Change

The phase analysis of simple speech utterances indicated that certain phase relations among the articulators remain unaltered across manifold speaker characteristics. Such critical phase angles are revealed by the flow of the dynamics of the system; they are not externally defined. As Sleigh and Barlow (1980) note in their comparative analysis of creatures that use a wide variety of propulsive structures for their activities, phase appears to provide essential information for stable coupling among the components of the system. How, then, do we conceive of the processes underlying *change* in articulatory pattern? What factors mediate the emergence of new (or different) spatiotemporal patterns? Such questions are at the heart of a theory of pattern generation. Below we offer an interpretation of certain kinds of articulatory (and phonetic) change in terms of the non-equilibrium phase transitions treated by synergetics (Haken, 1975, 1977; Haken et al., 1985). The central aspects of the theoretical model will be introduced briefly using an example from hand movements, and an application to speech will follow that focuses primarily on the effects of scaling changes in speaking rate, one of Stetson's (1928/1951) "great causes of phonetic modification" (p. 67). Importantly, the present analysis attests to the further significance of phase information--both within the stable *and* transition regions of the speech system's parameter space--in guaranteeing phonetic stability on the one hand and promoting phonetic change on the other. Moreover, along with other kinds of data (primarily on phonological development) the form of articulatory change may help rationalize a particular phonological unit, as a "natural" or intrinsically stable unit in the production of speech.

4.1 A Synergetic Outline: Pattern Formation and Change

For some years now, we have advocated an approach in which the control and coordination of multidegree of freedom speech and limb activities are treated in a manner continuous with cooperative phenomena in other physical, chemical, and biological systems, that is, as synergetic or dissipative structures (e.g., Kelso et al., 1981, 1983; Kelso & Saltzman, 1982; Kugler et al., 1980). These are systems--like that for speech production--that are composed of very many subsystems. In synergetics, when a certain parameter or combination of parameters (generally referred to as "controls") is scaled in sometimes quite nonspecific ways (i.e., the control prescription is not highly detailed), well-defined spatiotemporal patterns can form. The latter are maintained by a continuous flux of energy (or matter) through the system (e.g., Haken, 1975;

Yates, Marsh, & Iberall, 1972). Although there is pattern formation in the nonequilibrium phenomena treated by synergetics (e.g., the hexagonal forms produced in the Bénard convection instability, the transition from incoherent to coherent light waves in the laser, the oscillating waves and macroscopic patterns of various kinds of chemical reaction, etc.), there are, strictly speaking, no special mechanisms--like motor programs--that contain or represent the pattern before it appears (for further examples see Kelso & Tuller, 1984a).

How pattern formation occurs in these systems can be visualized roughly as follows. Imagine an open, dissipative system, one into which energy is continuously fed and from which it is continually dissipated. Certain configurations--called modes--are more capable of absorbing the energy flow than others. At a critical point, a linearized stability analysis reveals that the amplitude of these so-called unstable modes grows exponentially, whereas the other modes (the so-called "damped" modes) decay. In many nonequilibrium systems, close to critical (or bifurcation) points, the number of unstable modes can be shown to be much smaller than the number of stable, damped modes. In fact, the latter can be completely eliminated mathematically, according to Haken's so-called slaving principle, thereby allowing a tremendous reduction of the degrees of freedom. For example, in the laser (see Haken, 1975), a reduction from 10^{14} degrees of freedom to a single degree of freedom has been obtained.

More formally, the slaving principle states that the amplitudes of the damped modes can be expressed by means of a small set of "unstable" mode amplitudes (the so-called order parameters). The consequence is that all the damped modes follow the order parameters adiabatically, so that the behavior of the whole system is then governed by the order parameters alone (see Haken, 1977/1983, Chapter 7). Watching Bénard convection, for example, one is impressed how the total behavior--at a critical point--is completely captured by a macroscopic, modal action. The motions of the many microscopic, molecular components are completely irrelevant at this point: a low dimensional, macroscopic observable (the order parameter) specifies the system's evolving pattern.

However, identifying order parameters, even for many physical and chemical systems, is not always an easy matter. Certain guidelines do exist, however, which can be used to select viable candidates. A main one is that the order parameter changes much more slowly than the subsystems it is said to govern. Relative phase[6] fits this criterion quite well, since it is the phasing structure of many different activities that is preserved across scalar transformations (Section 3). Thus the individual articulatory components change quite a bit (kinematically and electromyographically), but the phase does not--at least in a given region of the parameter space.

4.2 Phase Transitions in Movement: An Explicit Example

Using relative phase as an order parameter, Haken, et al. (1985) have offered an explicit theoretical model of phase transitions that occur in bimanual activity (see Kelso 1981, 1984). The basic phenomenon is as follows: A human subject is asked to cycle his/her fingers at a preferred frequency using an out-of-phase, antisymmetrical motion. That is, flexion [extension] of one hand is accompanied by extension [flexion] of the other. Under an instruction to increase cycling frequency, that is, a systematic rate increase, the movements shift abruptly to an in-phase, symmetrical mode involving activation of homologous muscle groups. When the transition frequency was expressed in units of preferred frequency, the resulting dimensionless ratio or critical value was constant for all subjects but one (who was not naive and who purposely resisted the transition--although with certain energetic consequences, see Kelso, 1984). A frictional resistance to movement lowered both preferred and transition frequencies, but did not change the critical ratio (~1.33), suggesting the presence of an intrinsic invariant metric.

For present purposes, the main features of the bimanual experiments are: (1) the presence of only two stable phase (or "attractor") states between the hands (see also Yamanishi, Kawato, & Suzuki, 1980, for further evidence); (2) an abrupt transition from one attractor state to the other at a critical, intrinsically defined frequency; (3) beyond the transition, only one mode (the symmetrical one) is observed; and (4) when the driving frequency is reduced, the system does not return to its initially prepared state, that is, it remains in the basin of attraction for the symmetrical mode.

The theoretical strategy employed by Haken et al. (1985) to account for the foregoing findings may be worth noting. First, they specified a potential function corresponding to the layout of modal attractor states (i.e., the stable in-phase and out-of-phase patterns), and showed how that layout was altered as a control parameter (driving frequency) was scaled. From the behavior of the potential function, they then derived the equations of motion for each hand, and a nonlinear coupling structure between the hands. Analytic derivations and consequent numerical simulation revealed that if the model system was started, or "prepared" in the out-of-phase mode, and driving frequency was increased slowly, the system remained in that mode until the solution of the coupled equations of motion became unstable. At this point, a jump occurred and the only stable stationary solution produced by the system corresponded to the in-phase mode (see Haken et al., 1985, for more details). Ongoing theoretical (Schöner, Haken, & Kelso, 1986) and empirical (Kelso & Scholz, 1985) work has revealed that the nonlinear coupling strength as well as fluctuations (both intrinsically generated due to noise in system parameters and extrinsically generated due to an added random forcing function) play an important role in effecting the modal transitions between the hands. Thus, Kelso and Scholz (1985), in new experiments, have found both "critical slowing down" and enhanced fluctuations in order parameter behavior as the transition is approached. These predictions follow directly from the synergetic treatment of nonequilibrium phase transitions (see e.g., Haken, 1984; Haken et al., 1985; Schöner, Haken, & Kelso, 1986) and are simply not part of more conventional accounts of "switching" behavior based on motor programs (cf. Schmidt, 1982, p. 316) or central pattern generators (cf. Grillner, 1982, p. 224).

4.3 Stetson's (1951) Experiments

Let us now see how this view of spatiotemporal pattern formation and change may apply to speech production. To do this we draw initially on Stetson's (1951) work and offer a theoretical interpretation of his experiments that is consistent with synergetics. Then we mention some new (as yet preliminary) data of our own (Kelso, Munhall, Tuller, & Saltzman, in preparation) suggesting that certain kinds of phonetic change correspond directly to phase transitions among articulatory gestures.

Stetson (1951) recognized that "...the modification of the articulations is one of the most important aspects for study in experimental phonetics" (p. 67), and that scaling changes in speaking rate offered a window into the "various types of modification of the factors of the syllable, or the changing conditions that throw them together or force them apart" (p. 67). We also hypothesize--by analogy to our discussions above--that rate changes may: a) reveal the most stable modes of coordination of the articulatory system and, in turn b) that these stable modes may rationalize why one phonological form, the CV syllable, tends to be a universal feature of all languages (cf. Abercrombie, 1967; Bell, 1971; Clements & Keyser, 1983).

Consider first an example, discussed in some detail by Stetson (1951). A subject produces the CVC syllable "pup" repetitively. As speaking rate is gradually increased, Stetson describes the following changes: The syllables, "pup, pup..." at first distinct, come closer together. As rate increases. the arresting consonant of each syllable "doubles" with the releasing consonant of the next syllable. Thus the first change can be annotated as: "pup, pup".... → "pup-pup..." At still higher rates, according to Stetson, it becomes impossible to execute the prescribed number of consonants per second, and the arresting consonant of each syllable drops out. This second change can be referred to as "singling"; "pup-pup..." goes to "pu' pu '...."

Such changes induced by increasing speaking rate are brought about, in Stetson's words, by the tendency of movements "either to get into step or to drop out in order to simplify the coordination" (p. 71) and, relatedly, because of a "universal tendency to simplify by eliminating the arresting consonant" (p. 81). But why this particular tendency should prevail is unclear. On the one hand, elimination of the arresting consonant "simplifies" coordination; on the other, the process is dictated by maximum articulatory rates: "singling" must occur at rates of around 4.5 syllables per sec, because such rates in turn entail eight consonant movements per sec (Stetson, 1951).

"Simplification" as a function of maximum articulatory rate cannot be the whole story, of course. For example, often "singling" occurs at a rate as low as 2.5 syllables/sec. Also, the arresting consonant does not always drop out; often it is said to "fuse" with the releasing consonant (Stetson, 1951). Therein lies a potential clue. That is, it may be the phasing among component gestures--one with another--that is a central aspect of phonetic stability and change (cf. Section 3.0). For example, in his studies of the combination of abutting consonants, Stetson notes that the arresting consonant of one syllable (e.g., "p" in "sap") and the releasing consonant of the next (i.e., "s" in "sap"), "quickly overlap and soon become simultaneous ..... a striking illustration of the movements of speech to get in phase" (p. 78). Moreover, as rate decreases, the movements of arresting and releasing consonants "merely slide apart" (p. 78). Thus, there is a strong hint in Stetson's experiments and writings that under scaling influences of speaking rate, certain phase relations among gestures are more stable than others. For example, in all "phonetic coordinations" to use Stetson's phrase, there is a preferred relation between the releasing consonant and the beat stroke of the syllable. Namely, the releasing consonant never drops out. According to Stetson (1951), it retains its position because it coincides with the syllable's beat stroke.[7] In addition, compound consonants are said to be produced by the "sliding" of the two movements, e.g., the continuant labial "m" and the continuant lingual "s" in the syllable "mass" slide to form "sma." Stetson's descriptive language in this respect is almost prophetic of current formalisms: abutting consonants are "attracted" (p. 80) one to the other. As part of the tendency for movements to coincide, one consonant movement is delayed and the other advanced (cf. Stetson, 1951, p. 80).

Though phase was never explicitly measured in Stetson's work, and his account of phonetic change is largely posed within his "chest pulse" framework, there appears to be a strong linkage between his results and our previous discussion of hand movements. In particular, it seems possible that both may fall under the theoretical rubric of nonequilibrium phase transitions. Such a view is supported on at least two grounds. First, in an as yet quite limited data set, we have examined interarticulator phasing when subjects produced the vowel-consonant combination /ip/ at progressively faster rates (Kelso, et al., in preparation). A shift to the CV form /pi/ occurred at a given rate and was characterized by an abrupt change in the phase relation between glottal aperture and lip aperture. Second, it seems possible to reinterpret some of Stetson's own data on syllable durations when speaking rate is increased, as consistent with our previous theoretical discussion of phase transitions.

4.4. Phase Transitions in Speech: Some Direct Evidence

The design of the following experiment was extremely simple. Infrared light emitting diodes were placed on the subjects' lips and jaw, thus allowing us to obtain the trajectories of these articulators. Similarly, the opening and closing of the glottis was monitored by transillumination (e.g., Baer, Löfqvist, & McGarr, 1983). Similar to some of Stetson's (1951) work, the subject was invited to produce the syllable /ip/ at a slow speaking rate and then instructed simply to speed up in a step-like manner. A complete trial consisted of a series of repetitive syllables produced in a single breath. Typically, a trial lasted about 10 to 12 sec. An identical procedure was employed for the syllable /pi/. Subjects performed at least five trials per syllable. Although data collection and analysis are not yet completed (presently three subjects have been run), the data are quite clear thus far.

Trajectories over time of lip aperture (i.e., a single variable representing vertical distance between upper and lower lips[8] and glottal aperture are shown in Figure 7(1b) and 7(2b) for part of a representative trial for each utterance. In these subfigures, the vertical ticks denote the onset of lip opening (consonant release) and the occurrence of peak glottal opening (maximum vocal fold abduction). The corresponding relative phase between the lip and glottal aperture motions is shown in Figure 7(1a) and 7(2a). The movements shown were sampled at 200 Hz (for details of signal processing techniques, see Kay, Munhall, V.-Bateson, & Kelso, 1985).

In the case of both /ip/ and /pi/ it is quite obvious that the phase relation between lip opening onset and peak glottal opening is practically invariant, but different for the two syllables, over the range of speaking rates examined (approximately 1 to 5 syllables/sec). For /pi/, peak glottal opening lags the onset of oral opening by a constant amount, roughly 40-45 degrees. For /ip/ the two events are almost coincident, up to a speaking rate of approximately 4 syllables/sec. Then, a clear jump in phase occurs, practically within a single cycle, to the phasing pattern for /pi/. Note that like the hand movement data, the phase transition occurs at well below maximum syllable rates, at least for CV syllables. Again, both forms of coordination are quite stable below the critical region: only the coordinative mode characteristic of the CV, however, exists beyond the transition. Except for the quantitatively different phase relations observed, these speech data mimic the pattern of results observed in the bimanual data.

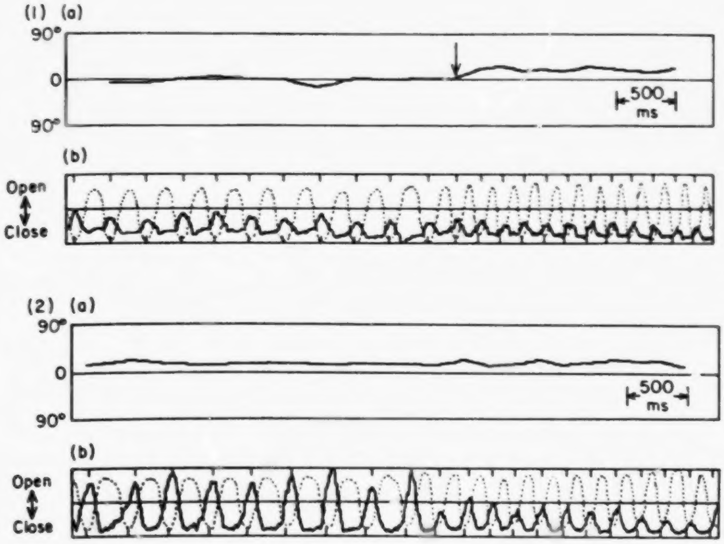


Figure 7. Phase relation between glottal and oral aperture (A), and trajectories of each variable over time (B) as the subject speeds up a given utterance. 1. The instructed utterance is /ip/. The arrow denotes a phase shift between glottal and oral aperture that occurs when the VC form /ip/ changes to the CV form, /pi/. 2. The instructed utterance is /pi/. Note that no phase shift occurs in this case even though the speaking rate is comparable to (1). See text for further details. ⟵——⟶ glottal aperture; (.....) lip aperture.

Several issues remain to be addressed, however. First, we need to know much more about what goes on in the region of the transition itself. Second, a continuous estimate of relative phase should be obtained (see Kelso & Scholz, 1985). The point estimate presented here requires the rather arbitrary selection of peaks or valleys in the time-series data as reference and/or target events. Given previous work on laryngeal-oral coordination (e.g., Löfqvist & Yoshioka, 1981), the selection of the present events (i.e., lip opening onset and peak glottal opening) seems reasonable. Obviously different events, for example, the onset of movement toward oral closure relative to peak glottal opening, would yield the same pattern but different phase values. A continuous estimate, based on a sample-by-sample phase difference, would not require one to make such a choice _a priori_. Third, Stetson (1951) reports that the original form restores with a decrease in rate, that is, the VC form returns and that "...this tendency to restoration ... is the great conservative factor in pronunciation" (p. 74). We suspect otherwise, though we have yet to check our suspicions formally. That is, once a transition to the CV form occurs, the system exhibits hysteresis--it tends to remain in the currently displayed form.[9] If this is so, then we have a model of phase transitions in speech that is formally equivalent to that of Haken, Kelso, and Bunz (1985) developed for phase transitions in hand movements.

4.5. Theory and Theoretical Implications

A further hint that the discontinuities created by rate scaling are at least consistent with a nonequilibrium phase transition interpretation can be gleaned from Stetson (1951, Figure 51). In this figure, whose main features are reproduced here (Figure 8a), distribution curves are presented showing the rates at which "doubling" and "singling" occur when a single syllable is repeated at varying rates. Although "doubling" occurs at rates between 1 and 3.5 syllables per sec, a peak for doubling is present at 2.5 syllables/sec. Another way to envisage these data is to invert the curves: Two minima are then apparent, one for each of the two articulatory patterns, separated by a local hill or maximum (Figure 8b). As speaking rate is increased, it becomes increasingly difficult (as indicated by progressively fewer and fewer observations) for a subject to maintain "doubling." Then, at a critical point, a shift to the next "equilibrium" configuration occurs, corresponding to the "singling" pattern. Analogous to our discussion above, it seems plausible to suggest that: a) the doubling and singling patterns correspond to distinct system modes, each characterized by specific phasing relations among articulatory gestures; and b) the transition from doubling to singling beyond a critical production rate reflects a system bifurcation.

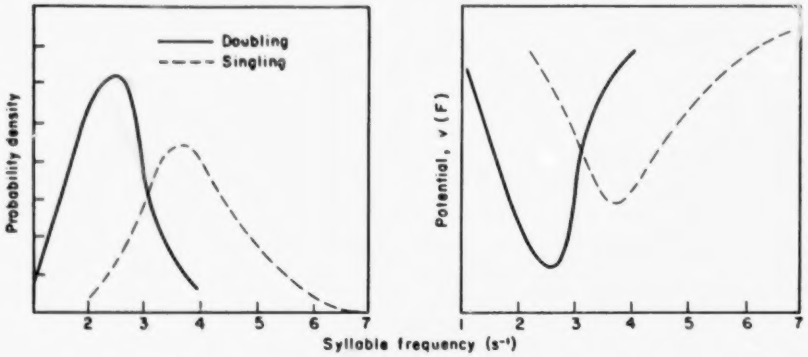


Figure 8. A. Distribution curves (probability density functions) showing the rates at which "doubling" and "singling" occur when a single syllable is repeated at varying syllable frequencies (adapted from Stetson, 1951, Figure 51, p. 69). B. Corresponding "potential functions" for the probability distributions shown in A. See text for further details.

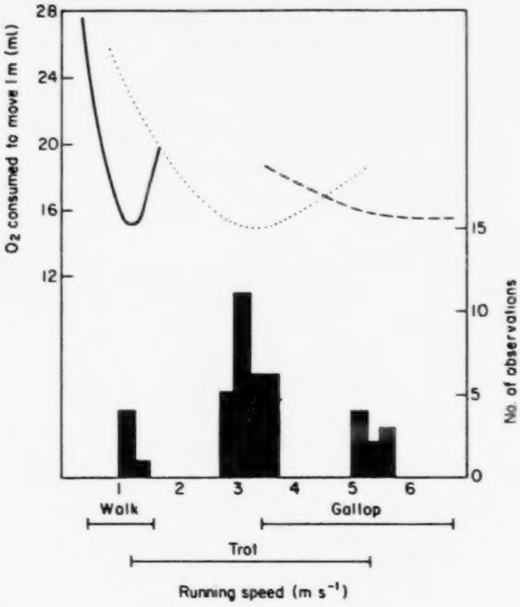


Figure 9. Relationship between oxygen consumption and locomotory speed during the walk, trot, and gallop of ponies (adapted from Hoyt & Taylor, 1981). See text for further details.

These speech data of Stetson, therefore, bear a striking resemblance to our speech and hand movement data as well as recent work on locomotor gait transitions (see Figure 9, reproduced from Hoyt & Taylor, 1981),[10] an interpretation of which is given in Kelso and Tuller (1984a) and Kelso and Scholz (1985). In the case of quadruped gait, the modes correspond to particular phasing relations among the limbs, which, when the animal is allowed to locomote freely, correspond to regions of minimum oxygen consumption. Hoyt and Taylor (1981), however, forced ponies to locomote away from these stability regions by increasing the speed of a treadmill on which the ponies walked. That is, according to our interpretation, they experimentally displaced the ponies away from equilibrium. As locomotor velocity is scaled, it becomes metabolically costly for the animal to maintain a given interlimb configuration; a switch into the next stable region, that is, the next local minimum, occurs (e.g., walking shifts to trotting). Like the hand movement data, when a critical value is reached (a point at which the "forces" driving through the system--roughly equated with increases in neural activation of muscle groups induced by rate scaling--compete with, and overcome the "forces" holding the system together, i.e., characteristic phase relations), the system bifurcates and a new (or different) spatiotemporal ordering emerges. We want to emphasize that such ordering changes are not strictly fixed for any of these situations. Horses, for example, can trot at speeds at which they normally gallop (as a visit to Yonkers' race track to observe the trotters will quickly reveal), but it is metabolically expensive to do so. Similarly "doubling" is possible beyond the bifurcation point in Figure 8, as illustrated by the dashed line, but there are so few observations there (at rates between 3 and 4 syllables/sec), as to suggest that coordination in that region is highly unstable.

In summary, although there are obvious differences between the various critical phenomena discussed here, there is reason to suppose that all of them--hand movement, speech, and gait--correspond to instabilities that arise as the particular system is driven experimentally away from equilibrium.[11] Obviously, much more work needs to be done to ground this conjecture (see Kelso & Scholz, 1985; Schöner et al., 1986, for possible experimental directions). In each case, new stabilities arise--indexed by particular phase relations between the components--as a result of competition between energy flowing into the operational components (i.e., a scaling influence) and the ability of those components to absorb the energy flow in their new configuration. In the hand movement case and, by hypothesis, in speech as well, we expect that higher bifurcations are possible because the system has available additional degrees of freedom (see Kelso & Scholz, 1985). That is, more configurations are possible--some of which will be stable and others not--precisely because of the availability of these extra degrees of freedom. In this view, the latter are not a curse (cf. Bellman, 1961) but a tremendous advantage. In addition, fluctuations (the "noise" often removed by engineers) can be shown to permit the discovery of new modes or phasing structures (cf. Schöner et al., 1986).

The present theoretical perspective not only affords the potential for a principled analysis of pattern formation (for many more details, see Haken et al., 1985), but, as we mentioned earlier, the nature of the pattern change itself may also prove rather informative. The theory predicts that the states that emerge under scaling influences are the most favored ones, and empirical evidence supports this interpretation. Thus, within the range of driving frequencies examined in the experiments of Kelso and colleagues, shifts to the symmetrical mode of coordination occur, but not vice-versa. Similarly, in speech, the articulatory configuration supporting the consonant-vowel form is the more fundamental: utterances never shift to the VC form under rate increases when the system is originally prepared in the "CV state." In each case, one structure can be fractured, but not the other. By the arguments and data discussed here, this is because certain phase relations among the articulators--which can be modeled as an order parameter for the total articulatory ensemble--are more stable than others.

Clearly both CV and VC forms (like symmetric and antisymmetric hand movements) can be produced easily in a given region of parameter space. The question of which of the two patterns is more basic, is answered by determining which remains beyond a critical point. The fact that the consonant-vowel form "wins out" when the system is scaled is thus a consequence of the stability of the articulatory configuration for that form. That is, certain configurations can absorb the energy input more efficiently than others. The universal tendency (Stetson, 1951) to simplify coordination by eliminating the arresting consonant (i.e., the one tied to the previous vowel), suggests that it is in some sense "easier" for the system to produce movements "in-phase," than otherwise. However, this does not have to be the case according to the present thesis. It remains very much an open question--to be pursued empirically--as to which phase relations are more stable than others. In the case of speech, unlike the hand movement case (at least in the most primitive, paradigmatic case studied in our experiments), we would expect a much larger and more varied (but perhaps nested) set of stable phasings. A similar hypothesis applies to studies of phasing in skilled pianists, which we are presently analyzing. In each case, the layout of the attractor states should be much more "wrinkled" than the "simple" bistable potential--differentiating in-phase and antiphase modes--that we have studied thus far.

In spite of the foregoing caveats, the present analysis may rationalize, in an elegant fashion, why the consonant-vowel is a core syllable type in all languages (cf. Abercrombie, 1967; Clements & Keyser, 1983). Such a rationale has been missing in much phonological theory, which starts off with the CV core unit as a basic assumption. Moreover, the developmental evidence reviewed by Locke (1983) reveals a strong tendency for consonant initial (CV) forms to predominate in infant babbling. Rate scaling studies may reveal these primitive forms of coordination in the mature organism and therefore offer a window into the building blocks of language, a principled decomposition of which has been lacking. Like the particle accelerator that breaks atoms apart to reveal their secrets, so forcing the articulatory system to perform at unusual rates may reveal the primitive units of language, and, more important, their interactions with other units. Lest this image be interpreted as too mechanical or immutable, let us allay the reader's concern; nothing could be further from our intent. Just as the cheetah does not have to proceed through the locomotory gaits when it pursues its prey, so the system that realizes language does not have to traverse through any fixed set of phase relations to reveal its intent. What this experimental program may reveal, and this theoretical framework rationalize, is a design that allows for, rather exploits, the low energy switching among its articulatory configurations--in short, a design appropriate for intentional systems.

5. Summary

Presented here, in preliminary form, is a general theoretical framework that seeks to characterize the lawful regularities in articulatory pattern that occur when people speak. A fundamental construct of the framework is the coordinative structure, an ensemble of articulators that functions cooperatively as a single task-specific unit. Direct evidence for coordinative structures in speech is presented and a control scheme that realizes both the contextually-varying and invariant character of their operation is outlined. Importantly, the space-time behavior of a given articulatory gesture is viewed as the outcome of the system's dynamic parameterization, and the orchestration among gestures is captured in terms of intergestural phase information. Thus, both time and timing are deemed to be intrinsic consequences of the system's dynamical organization. The implications of this analysis for certain theoretical issues in coarticulation raised by Fowler (1980) receive a speculative, but empirically testable, treatment. Building on the existence of phase stabilities in speech and other biologically significant activities, we also offer an account of change in articulatory patterns that is based on the nonequilibrium phase transitions treated by the field of synergetics. Rate scaling studies in speech and bimanual activities are shown to be consistent with a synergetic interpretation and suggest a principled decomposition of languages. The CV syllable, for example, is observed to represent a stable articulatory configuration in space-time, a possible rationalization for the presence of the CV as a phonological form in all languages. The uniqueness of the present scheme is that stability and change of speech action patterns are seen as different manifestations of the same underlying dynamical principles--the phenomenon observed depends on which region of the parameter space the system occupies. Though probably wrong, ambitious, and the outcome of much idle speculation, the simplicity of the present scheme is attractive and may offer certain unifying themes for the traditionally disparate disciplines of linguistics, phonetics, and speech motor control.

References

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior*, *16*, 195-231.

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.

Baer, T., Löfqvist, A., & McGarr, N. S. (1983). Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. *Journal of the Acoustical Society of America*, *73*, 1304-1308.

Bell, A. (1971). Some patterns of occurrence and formation of syllable structures. *Stanford University, Department of Linguistics, Working Papers on Language Universals*, *6*, 23-137.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. *Phonetica*, *38*, 9-20.

Bellman, R. E. (1961). *Adaptive control processes: A guided tour*. Princeton, NJ: Princeton University Press.

Benati, M., Gaglio, S., Morasso, P., Tagliasco, V., & Zaccaria, R. (1980). Anthropomorphic robotics. II. Analysis of manipulator dynamics and the output motor impedance. *Biological Cybernetics*, *38*, 141-150.

Bernstein, N. A. (1967). *The coordination and regulation of movements*. London: Pergamon Press.

Berkenblit, M. B., Fel'dman, A. G., & Fukson, O. I. (in press). Adaptability of innate motor patterns and motor control mechanisms. *The Behavioral and Brain Sciences*.

Browman, C. P., & Goldstein, L. (in press). Towards an articulatory phonology. *Phonology Yearbook*.

Browman, C. P., Goldstein, L., Kelso, J. A. S., Rubin, P., & Saltzman, E. (1984). Articulatory synthesis from underlying dynamics. *Journal of the Acoustical Society of America*, *75*, S22-S23. (Abstract)

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1984). Inadequacies of the computer metaphor. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum.

Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Coker, C. H. (1976). A model of articulatory dynamics and control. *Proceedings of the IEEE*, *64*, 452-460.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North Holland.

Edelman, G. M., & Mountcastle, V. B. (1978). *The mindful brain*. Cambridge, MA: MIT Press.

Fowler, C. (1977). *Timing control in speech production.* Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics*, *8*, 113-133.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press.

Fukson, O. I., Berkinblit, M. B., & Fel'dman, A. G. (1980). The spinal frog takes into account the scheme of its body during the wiping reflex. *Science*, *209*, 1261-1263.

Gentil, M., Harris, K. S., Horiguchi, S., & Honda, K. (1984). Temporal organization of muscle activity in simple disyllables. *Journal of the Acoustical Society of America*, *75*, S23. (Abstract)

Greene, P. H. (1971). Introduction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), *Models of the structural-functional organization of certain biological systems*. Cambridge, MA: MIT Press.

Grillner, S. (1977). On the neural control of movement--A comparison of different basic rhythmic behaviors. In G. S. Stent (Ed.), *Function and formation of neural systems*. Berlin: Dahlem.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control*. Oxford: Pergamon Press.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. *Review of Modern Physics*, *47*, 67-121.

Haken, H. (1977). *Synergetics: An introduction. Nonequilibrium phase transitions and self-organization in physics, chemistry, and biology* (Third edition, 1983). Heidelberg: Springer Verlag.

Haken, H. (1983). *Advanced synergetics*. Heidelberg: Springer-Verlag.

Haken, H. (1984). Synergetics. *Physica*, *127B*, 26-36.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347-356.

Harris, K. S., Tuller, B., & Kelso, J. A. S. (1986). Temporal invariance in the production of speech. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes*. Hillsdale, NJ: Erlbaum.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Hollerbach, J. (1981). An oscillator theory of handwriting. *Biological Cybernetics*, *39*, 139-156.

Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. *Nature*, *292*, 239-240.

Kay, B. A., Munhall, K. G., V.-Bateson, E., & Kelso, J. A. S. (1985). A note on processing kinematic data: Sampling, filtering, and differentiation. *Haskins Laboratories Status Report on Speech Research*, *SR-81*, 291-303.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 529-543.

Kelso, J. A. S. (1978). Joint receptors do not provide a satisfactory basis for motor timing and positioning. *Psychological Review*, *85*, 474-481.

Kelso, J. A. S. (1981). On the oscillatory basis of movement. *Bulletin of the Psychonomic Society*, *18*, 63.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, *15*, R1000-R1004.

Kelso, J. A. S. (in press). Pattern formation in multidegree of freedom speech and limb movements. *Experimental Brain Research Supplementum*.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. *Journal of Motor Behavior*, *13*, 226-261.

Kelso, J. A. S., Munhall, K. G., Tuller, B., & Saltzman, E. L. (in preparation). Manuscript in preparation.

Kelso, J. A. S., & Saltzman, E. (1982). Motor control: Which themes do we orchestrate? *Behavioral and Brain Sciences*, *5*, 554-557.

Kelso, J. A. S., & Scholz, J. (1985) Cooperative phenomena in biological motion. In H. Haken (Ed.), Complex systems: Operational approaches in neurobiology, physical systems, and computers. Berlin: Springer-Verlag.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Kelso, J. A. S., & Tuller, B. (1984a). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience (pp. 321-356). New York: Plenum.

Kelso, J. A. S., & Tuller, B. (1984b). Converging evidence in support of common dynamical principles for speech and movement coordination. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 246, R928-R935.

Kelso, J. A. S., & Tuller, B. (1985a). Intrinsic time in speech production: Theory, methodology, and preliminary observations. Haskins Laboratories Status Report on Speech Research, SR-81, 23-39. Also (in press) in E. Keller and M. Gopnik (Eds.), Sensory and motor processes in language, Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Tuller, B. (1985b). Intrinsic time in speech production. Journal of the Acoustical Society of America, 77(Suppl.1), S53.

Kelso, J. A. S., Tuller, B., & Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. Journal of the Acoustical Socety of America, 72, S103.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech (pp. 137-173). New York: Springer-Verlag.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1986). A theoretical note on speech timing. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., Tuller, B., V.-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Klein, C. A., & Huang, C.-H. (1983). Review of pseudoinverse control for use with kinematically redundant manipulators. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13, 245-250.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of movement control and coordination (pp. 5-78). Chichester: John Wiley.

Ladefoged, P., Draper, A., & Whitteridge, P. (1958). Syllables and stress. Miscellanea Phonetica, 3, 1-14.

Lestienne, F. (1979). Effects of inertial load and velocity on the braking process of voluntary limb movements. Experimental Brain Research, 35, 407-418.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory Quarterly Progress Status Report*, *STL-QPSR-4*, 1-29.

Linville, R. (1982). *Temporal aspects of articulation: Some implications for speech motor control of stereotyped productions*. Unpublished doctoral dissertation, University of Iowa.

Locke, J. (1983). *Phonological acquisition and change*. New York: Academic Press.

Löfqvist, A., & Yoshioka, H. (1981). Interarticulator programming in obstruent production. *Phonetica*, *38*, 21-34.

Lubker, J. (1983, October). *Comment on "Temporal invariance in the production of speech," by Harris, Tuller, and Kelso*. Paper presented at Conference on invariance and variability in speech processes. MIT, Cambridge, MA.

MacNeilage, P. F., Hanson, R., & Krones, T. (1970). Control of the jaw in relation to stress in English. *Journal of the Acoustical Society of America*, *48*, 120(A).

Maxwell, J. C. (1877). *Matter and motion*. New York: Dover Press.

Munhall, K. G. (in press). An examination of intra-articulator relative timing. *Journal of the Acoustical Society of America*.

Munhall, K. G., & Kelso, J. A. S. (1985, November). *Phase dependent sensitivity to perturbation reveals the nature of speech coordinative structures*. Paper presented at the meeting of the Acoustical Society of America, Nashville, TN.

Nashner, L. M. (1977). Fixed patterns of rapid postural responses among leg muscles during stance. *Experimental Brain Research*, *30*, 13-24.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. *Science*, *201*, 1235-1237.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos*. New York: Bantam Books.

Rosen, R. (1980). Protein folding: A prototype for control of complex systems. *International Journal of Systems Science*, *11*, 527-540.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, *70*, 321-328.

Saltzman, E. (1979). Levels of sensorimotor representation. *Journal of Mathematical Psychology*, *20*, 91-163.

Saltzman, E. (1985). Task dynamic coordination of the speech articulators: A preliminary model. *Haskins Laboratories Status Report on Speech Research*, SR-84, 1-18. Also *Experimental Brain Research Supplementum*, in press.

Saltzman, E. L., & Kelso, J. A. S. (1983). Skilled actions: A task dynamic approach. *Haskins Laboratories Status Report on Speech Research*, *SR-76*, 3-50. Also *Psychological Review*, in press.

Schmidt, R. A. (1982). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics.

Schmidt, R. A., & McGown, C. (1980). Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. *Journal of Motor Behavior*, *12*, 149-161.

Schöner, G., Haken, H., & Kelso, J. A. S. (1986). A stochastic theory of phase transitions in human hand movement. *Biological Cybernetics*.

Shapiro, D. C., Zernicke, R. F., Gregor, R. J., & Diestal, J. D. (1981). Evidence for generalized motor programs using gait-pattern analysis. Journal of Motor Behavior, 13, 33-47.

Shik, M. L., & Orlovskii, G. N. (1965). Coordination of the limbs during running of the dog. Biophysics, 10, 1148-1159.

Sleigh, M. A., & Barlow, D. I. (1980). Metachronism and control of locomotion in animals with many propulsive structures. In H. Y. Elder & E. T. Trueman (Eds.), Aspects of animal locomotion. Cambridge: Cambridge University Press

Stetson, R. H. (1951). Motor phonetics: A study of speech movements in action. Amsterdam: North-Holland.

Stone, M. (1981). Evidence for a rhythm pattern in speech production: Observations of jaw movement. Journal of Phonetics, 9, 109-120.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. Journal of Speech and Hearing Research, 16, 397-420.

Tuller, B., Harris, K., & Kelso, J. A. S. (1982). Stress and rate: Differential transformations of articulation. Journal of the Acoustical Society of America, 71, 1534-1543.

Tuller, B., & Kelso, J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariants. Journal of the Acoustical Society of America, 76, 1030-1036.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 8, 460-472.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1983). Further evidence for the role of relative timing in speech: A reply to Barry. Journal of Experimental Psychology: Human Perception and Performance, 9, 829-833.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum.

Viviani, P., & Terzuolo, V. (1980). Space-time invariance in learned motor skills. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Whitney, D. E. (1972). The mathematics of coordinated control of prosthetic arms and manipulators. ASME Journal of Dynamic Systems, Measurement and Control, 94, 303-309.

Yamanishi, J., Kawato, M., & Suzuki, R. (1980). Two coupled oscillators as a model for coordinated finger tapping of both hands. Biological Cybernetics, 37, 219-225.

Yates, F. E., Marsh, D. J., & Iberall, A. S. (1972). Integration of the whole organism: A foundation for a theoretical biology. In J. A. Behnke (Ed.), Challenging biological problems: Directions towards their solution (pp. 110-132). New York: Oxford University Press.

Footnotes

[1] Note that this claim, to be supported here, seems to run counter to the notion that normal phoneme rates "can be achieved only if separate parts of the articulatory machinery--muscles of the lips, tongue, velum, etc.--can be separately controlled" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

[2] With implications, no doubt, for speech perception, a topic that is, however beyond the mission of this paper (but see, e.g., Liberman & Mattingly, 1985).

[3] A preliminary report of these data was given by Kelso and Tuller (1985b).

[4] Note that there is an important caveat for the phase notion. Though phase has been illustrated here at a very "simple" interarticulator level, we do not want to suggest that this is necessarily the appropriate frame of reference for speech production and perception. However, the point is that regardless of the particular frame of reference (e.g., events defined at muscle, articulator, tract variable levels, etc.), a concept such as phase will be crucial to specifying the sequence of events.

[5] The computation of the phase angle of a point on the phase plane is problematical. Because the units of position and velocity are incommensurate (mm vs. mm/sec, for example), applying inverse trigonometric functions directly to the data yields meaningless results. To avoid this problem, we normalize both position and velocity to the same numerical interval, -1 to +1 (not necessarily the unit circle for periodic data), and then apply the inverse tangent function to the normalized data. The normalization of position over a cycle of data proceeds via the following linear transform:

$$P_{norm} = 2P/(P_{max}-P_{min}) - (P_{max}+P_{min})/(P_{max}-P_{min}), \quad (1)$$

where P_{norm} is the normalized position, P is the actual position, and P_{max} and P_{min} are the maximum and minimum position values over a cycle.

This has the effect of (i) rescaling the data to a range of 2 units and (ii) shifting the equilibrium position to zero, i.e., the interval -1 to +1 is achieved. Velocity is normalized according to which half-cycle the point of interest is in, to put the half-cycle of articulator raising on the interval 0 to 1 and the half-cycle of articulator lowering on the interval 0 to -1. That is,

$$V_{norm} = V/abs(V_{max}), \quad (2)$$

where V_{norm} is the normalized velocity, V is the actual velocity, and V_{max} is the maximum velocity during the corresponding half-cycle.

The arctangent was then computed using the normalized position and velocity values,

$$\text{phase angle} = \arctan(V_{norm}/P_{norm}) \quad (3)$$

The final value obtained is a number from 0 to 360 degrees, which increases in value in the direction opposite to the unwinding of trajectories on the phase plane, as per mathematical convention.

[6] A key feature in the development of science has been to define limits or constraints on natural phenomena. Once such constraints are known, much new understanding results (Prigogine & Stengers, 1984). Phase has this constraint-like property. Our experiments described in Section 4 reveal the limits over which one organization (a given phase relation) can remain stable. Also, because in those experiments, it is phase (and phase alone, as far as we know) that changes dramatically, we have some reason to suppose that phase is a key parameter even in the stable range of performance, that is, that phase represents a fundamental constraint (see Section 3).

[7] The beat stroke as defined by Stetson (1951), is "always ballistic ... and can hardly be longer than 40-100 ms" (p. 29). He continues "The unit movement of speech is the pulse which produces the syllable, a pulse of air through the glottis made audible by the vocal folds in speaking aloud and stopped and started by the chest muscles or by the auxiliary movements of the consonants" (p. 30). And later on in a discussion of consonant release, Stetson indicates that "... the stroke of the expiratory chest muscles and the beat stroke of the consonant occur at the same time" (p. 46). We include this definition and clarification of "beat stroke" for mostly historical reasons. Some of Stetson's claims about syllable pulses have been seriously questioned (Ladefoged, Draper, & Whitteridge, 1958). The present analysis, of course, does not rely on such notions.

[8] Lip aperture was estimated simply by subtracting the position of the lower lip from that of the upper lip. Note, however, that the same data pattern was obtained when the movement of a single labial articulator (e.g., the lower lip) was compared to glottal aperture.

[9] Stetson (1951) says little about the instructions given to subjects when speaking rate is reduced. Obviously, with slowing they will be able to say /ip/ below a certain rate. The question is when they do so spontaneously if, for example, they could not hear the consequences of their production.

[10] This resemblance is not only qualitative but perhaps quantitative as well. It may be pure serendipity that the ratio of the "doubling" mode frequency (~2.5 syllables/sec) to the critical frequency (~3.1 syllables/sec), shown in Figure 8, bears a close correspondence (~1.24) to the dimensionless ratios computed for Kelso's bimanual (~1.31) and Hoyt and Taylor's gait (~1.33) data (see Kelso, 1984). These dimensionless numbers, analogous to Reynolds' numbers in fluid dynamics, may be a reflection of the system's intrinsic "distance from equilibrium." That is, they may index how far beyond a "preferred" steady state a given pattern can persist before it fractures into a new configuration.

[11] One difference, for example, between the hand and speech data and the gait analysis is that, in the former, the various modal patterns can coexist in stable forms at subcritical rates. Galloping, on the other hand, is not observed at slow walking speeds. Though it may be available, it is simply not a stable locomotor mode in that region of the parameter space (see Figure 9).

THE VELOTRACE: A DEVICE FOR MONITORING VELAR POSITION*

Satoshi Horiguchi† and Fredericka Bell-Berti††

Abstract. This paper describes the Velotrace, a mechanical device designed to allow the collection of analog data on velar position. The device consists of two levers connected through a push-rod and carried on a pair of thin support rods. The device is positioned in the nasal passage with the internal lever resting on the nasal surface of the velum and the external lever positioned outside the nose. The movements of the external lever reflect the movement of the internal lever as it follows velar movement and are recorded as an analog signal using an optoelectronic position-sensing system. Results of evaluation studies indicate that the Velotrace accurately reflects the relatively rapid movements of the velum during speech.

Introduction

Since the size of the velar port determines the oral or nasal nature of speech sounds, there has long been interest in studying the velopharyngeal region (see Fritzell, 1969, for an extensive historical review). The various techniques used to study the velopharyngeal mechanism have examined a number of its dimensions, one result of which is the recognition that the size of the open velar port is reflected in the position of the velum, although velar position may also vary when the port is completely closed (see, for example, Henderson, 1984; Moll & Daniloff, 1971). These adjustments of velar position above the level at which closure occurs result from the anatomical relationship between the velum and the levator veli palatini (LVP) muscle. That is, since the superior attachment of the LVP muscle lies well above the level at which port closure is complete, increasing contraction of that muscle will continue to raise the velum even after the velopharyngeal port has been closed. As a result, changes in the vertical position of the velum throughout its range of movement may be considered to reflect speech motor control of the velum, and have the additional benefit of not suffering from a boundary effect in the way that velar port size measures do when port closure is achieved (see, for example, Bell-Berti, 1980). Thus, monitoring changes in the vertical position of the velum should allow the discovery of the principles of the (normal) velar motor control, which should increase our understanding of speech production, in general, and also increase our ability to evaluate velar

---

*Versions of the paper were presented at the 1984 meeting of the American Cleft Palate Association (Seattle, WA, May 1984) and of the American Speech-Language-Hearing Association (San Francisco, CA, November 1984).

†Also Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo, Tokyo, Japan

††Also St. John's University.

Acknowledgment. This work was supported by NINCDS grants NS-13617 and NS-13870 and BRSG grant RR-05596 to the Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-85 (1986)]

control problems in some clinical populations. Thus, and continuing in the tradition of mid-sagittal monitoring of velar function, we have developed a new mechanical device, the Velotrace, that allows the collection of data on velar position in analog form and eliminates the need for X-ray exposure and for frame-by-frame measurements of cine and video recordings.

The Device

The Velotrace (Figure 1) has three major parts: an internal lever, an external lever, and a push-rod between them. The push-rod and levers are carried on a pair of thin support rods. The levers are so connected to the push-rod that when the internal lever is raised, the external lever is deflected toward the subject. The device is loaded with a small spring that improves its frequency response and thus improves the ability of the internal lever to follow rapid downward movements of the velum. The effective length of the internal lever is 30mm (i.e., the linear distance between the fulcrum and tip), that of the external lever is 60mm, and that of the push-rod assembly is 150mm. The height of the device is 4mm and its width is 3mm, making it no larger than many commonly used nasopharyngeal fiberoptic endoscopes.

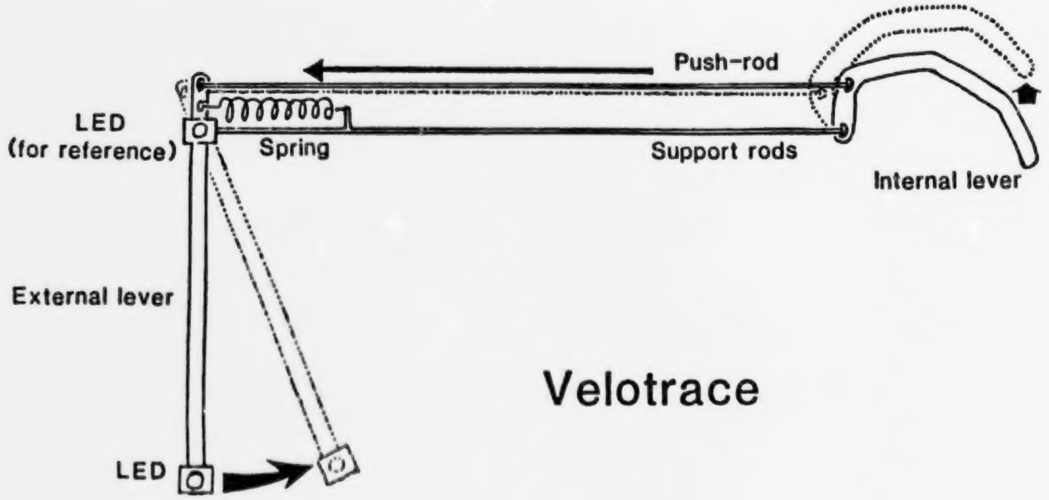


Figure 1. Schematic diagram of the Velotrace.

The device is positioned after topical anesthetic and decongestants have been applied to the nasal mucosa, if necessary, and the posterior pharyngeal wall has become visible through the nasal passage; the Velotrace is inserted using a procedure similar to that used for nasal catheterization. Although the Velotrace is a rigid device (unlike fiberoptic endoscopes), the insertion is easy unless the subject has serious pathologies or deformities in the nasal passage (e.g., substantial deviation of the nasal septum, nasal polyps, etc.).

None of our four subjects (three males and one female) for the evaluation study has complained of any discomfort from the device.

The fulcrum of the internal lever of the Velotrace is positioned at the end of the hard palate, with the internal lever resting on the velum and the support rods resting on the floor of the nasal cavity (Figure 2). An external clamp, which is attached to a headband positioned on the subject's head, is used to stabilize the position of the Velotrace against his/her head during recording sessions.

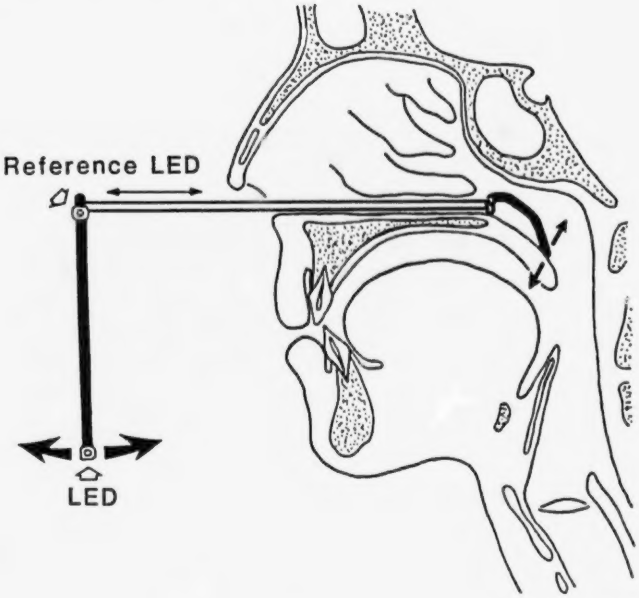


Figure 2. A mid-sagittal schematic drawing of the Velotrace in position with the internal lever resting on the velum.

The Recording System

Monitoring the movements of the external lever can be accomplished in a number of different ways. For example, one might use a velocity-displacement transducer, which would make the Velotrace a convenient stand-alone device for the clinical evaluation of velar movement. Another approach would be to use an optoelectronic tracking system, such as the one we have been using, to monitor the movements of the external lever using infrared Light Emitting Diodes (LEDs) attached to the Velotrace. In our system, one LED is attached to the end of the external lever and allows us to monitor the movement of the lever about its fulcrum. A second LED is positioned at the fulcrum of the external lever and serves as a reference point against which the movements of the end of the external lever can be described. The positions of the LEDs are tracked in two-dimensional space. The acoustic speech signal and a timing signal are recorded simultaneously with the LED-position signals on a multi-channel instrumentation data recorder. The position signals may also be monitored with an oscilloscope in real time. The data acquisition system is represented in Figure 3.

Evaluation Studies

The experimental utterance set was composed of three groups of disyllables (Table 1). The eight items of Utterance Group 1 each contained a medial oral-nasal consonant contrast that was used to insure that maximum stress was placed on the velar lowering mechanism (because the nasal consonant immediately follows a very strongly oral articulation). These utterances allowed us to examine the ability of the Velotrace to follow very rapid downward movements of the velum. Conversely, the eight items of Utterance Group 2 each contained a medial nasal-oral consonant contrast, used to insure that maximum stress was placed on the velar raising mechanism (because a very strongly oral articulation immediately follows a nasal one). These utterances allowed us to examine the ability of the Velotrace to follow very rapid upward movements of the velum. The six items of Utterance Group 3 contained high and low vowel contrasts with medial oral consonant sequences of varying length, and allowed us to examine the ability of the Velotrace to reflect the smaller velar excursions of entirely oral speech. All of the utterances used also had the advantage of having been used in some of our previous work, thus providing the opportunity of comparing the Velotrace data, albeit for different subjects, with endoscopically recorded data.

In the first evaluation study, Velotrace data were compared with previously collected endoscopic data. The endoscopic data used for comparison with the Velotrace data were obtained from two experiments in which frame-by-frame measurements of velar position were made of cine films photographed through a nasally positioned fiberoptic endoscope (Bell-Berti, 1980; Bell-Berti, Baer, Harris, & Niimi, 1979). The subject for the endoscopic studies was a speaker of educated Greater Metropolitan New York City English. In those experiments a long thin plastic strip with grid markings was inserted into the subject's nostril and placed along the floor of the nose and over the nasal surface of the velum, to enhance the contrast between the edge of the supravelar surface and the posterior pharyngeal wall. Then a flexible fiberoptic endoscope was inserted into the subject's nostril, and positioned so that it rested on the floor of the nasal cavity with its objective lens at the posterior border of the hard palate, providing a view of the velum and lateral pharyngeal walls from the level of the hard palate to above the maximum elevation of the velum (observed during blowing). Cine films were taken through the endoscope at 60 frames/sec. The position of the high point of the velum was then tracked, frame-by-frame, with the aid of a small laboratory computer.

The subject for the first Velotrace experiment was a normal speaker of educated Middle Atlantic American English who produced between 7 and 12 repetitions of each of the 22 experimental utterances. The 16 disyllables of Groups 1 and 2 were produced in isolation. The six disyllables of Group 3 were produced in the carrier phrase "It's a (test word) again." Within each group, the utterances were read from randomized lists. The speech acoustic signal and the positions of the LEDs were recorded simultaneously using the system described above. The speech acoustic signal was subsequently digitized at 10,000 samples/sec and the Velotrace signals were digitized at 200 samples/sec.

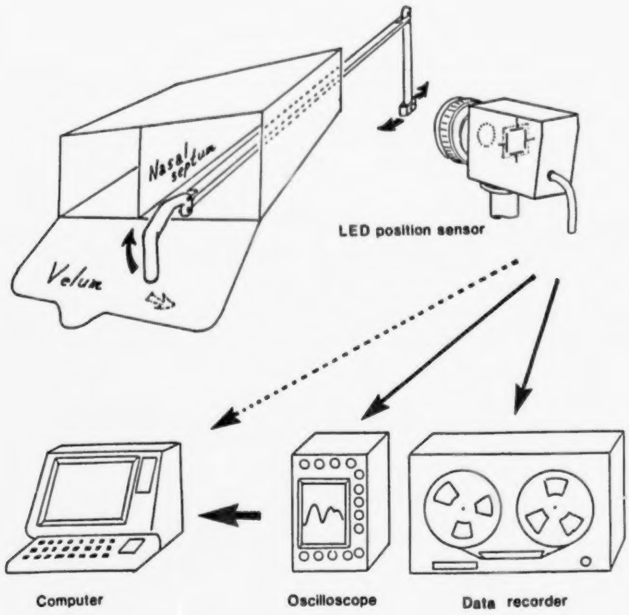

Nasal septum
Velum
LED position sensor
Computer
Oscilloscope
Data recorder

Figure 3. Schematic diagram of the recording and processing system.

Table 1

Experimental Utterance List

Utterance Group 1 (oral-nasal contrast)	Utterance Group 2 (nasal-oral contrast)	Utterance Group 3 (vowel contrast)
/fipmip/	/fimpip/	/flisɑp/
/fɑpmɑp/	/fɑmpɑp/	/flitsɑp/
/fibmip/	/fimbip/	/flistѕtɑp/
/fɑbmɑp/	/fɑmbɑp/	/kɑsiz/
/fismip/	/fimsip/	/kɑtsiz/
/fɑsmɑp/	/fɑmsɑp/	/kɑststiz/
/fizmip/	/fimzip/	
/fɑzmɑp/	/fɑmzɑp/	

An acoustic event identified in the waveform of each token of each utterance type served as a reference point for that token in subsequent data analysis. The choice of acoustic reference point depended upon the phonetic structure of the utterance.[1] These reference points have two functions: First, they allow us to examine the physiological signals for repetitions of an utterance type with reference to the same acoustic event. Second, they provide a reference point for aligning tokens of an utterance for calculating an ensemble average of the signals for the repetitions of an utterance type.

The endoscopically collected velar position data had been reduced to ensemble averages and their standard deviations (Bell-Berti, 1980; Bell-Berti et al., 1979). Since the individual token data were no longer available, it was necessary to calculate the equivalent ensemble averages for the Velotrace data. However, before comparing ensemble averages of the Velotrace data with the endoscopic data, we examined the Velotrace token data and averaged data. Samples of both ensemble-average and individual token Velotrace data are shown in Figure 4. The velar movement patterns recorded with the Velotrace are very similar for the tokens of each utterance type, and the individual tokens are also strikingly similar to the ensemble averages. Thus, we conclude that the ensemble averages are representative of the constituent token data, and may be used for comparison of Velotrace data with existing ensemble-average endoscopic data.

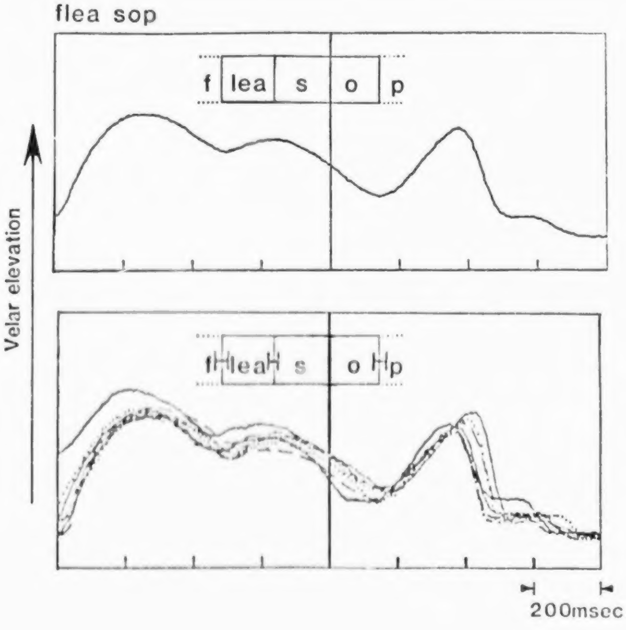


Figure 4. Ensemble-average (above) and individual token (below) Velotrace data for one utterance type. Zero on the abscissa identifies the reference point for aligning tokens of an utterance type for computer sampling and averaging.

Figure 5 displays ensemble averages of two different test words, (one each from Groups 1 and 2) recorded with an endoscope for one subject and with the Velotrace for the other subject. It is clear that the ensemble-averaged Velotrace data display the same patterns as do the frame-by-frame measurement data obtained from the cine films, although the subject, speech rate, and duration of the individual speech sounds are different. We also observe strikingly similar patterns for endoscopic and Velotrace data in which the test words were embedded in a carrier phrase (Figure 6). (See Bell-Berti, 1980, for a description of the experimental design and results of the second endoscopic study.)

For the second evaluation study, cine-radiographic films were taken of a third subject, also a speaker of educated Greater Metropolitan New York City English. The experimental utterances were two tokens each of a subset of six of the utterances used by Bell-Berti et al. (1979) and in the first evaluation study. For this experiment, the Velotrace was positioned in the subject's nasal cavity, with the internal lever resting on the nasal surface of the velum. A thin gold chain was inserted through the other nasal passage and positioned along the velum and into the oropharynx to improve visualization of the nasal surface of the velum in the X-ray images.[2] The films were taken at 60 frames/sec.

Figure 7 represents the film frame image, with the measurement points indicated with numbers on the figure. We measured the position of the tip of the internal lever of the Velotrace (1), the point on the velum that would be tracked by the Velotrace (2), the Velotrace internal fulcrum (3), and two reference points: an upper molar (4) and a lead pellet on the upper incisor (5). Visual inspection of the data on the vertical position of the Velotrace lever and of the velum (see Figure 8) suggests that movements of the Velotrace clearly reflect the movements of the velum itself. In order to quantify the relationship between these measures, we calculated the correlation coefficient between our measures for each of the twelve tokens. The very high linear correlation between these two measurements is reflected in scatterplots of the data (e.g., Figure 9) and in correlation coefficients of between 0.982 and 0.995.

We also compared velocity measures derived from these Velotrace movement data with equivalent velocity measures reported in the literature. To do this, we calculated the velocity of the vertical component of the velar and Velotrace movements. The velocity functions were calculated from successive central difference scores for each sample point. Maximum upward velocity, occurring in the transition between a nasal and a fricative consonant (/fimzip/), may be as high as 130mm/sec; maximum downward velocity, occurring in the transition between fricative and nasal consonants (/fismip/), may be as high as 100mm/sec. These values are similar, but not identical to Kuehn's (1976) data, in which he reports downward velocities as high as 132mm/sec.[3] We also calculated the linear correlation between the velar and Velotrace velocity functions for each token; all are within the range of r=0.90 to r=0.97. On the basis of visual inspection of the functions, the very high positive linear correlations between the two positional measures for each of the 12 tokens in our experimental set, and the equally high correlation between the velocities of these movements, we conclude that the internal lever accurately follows movements of the velum.

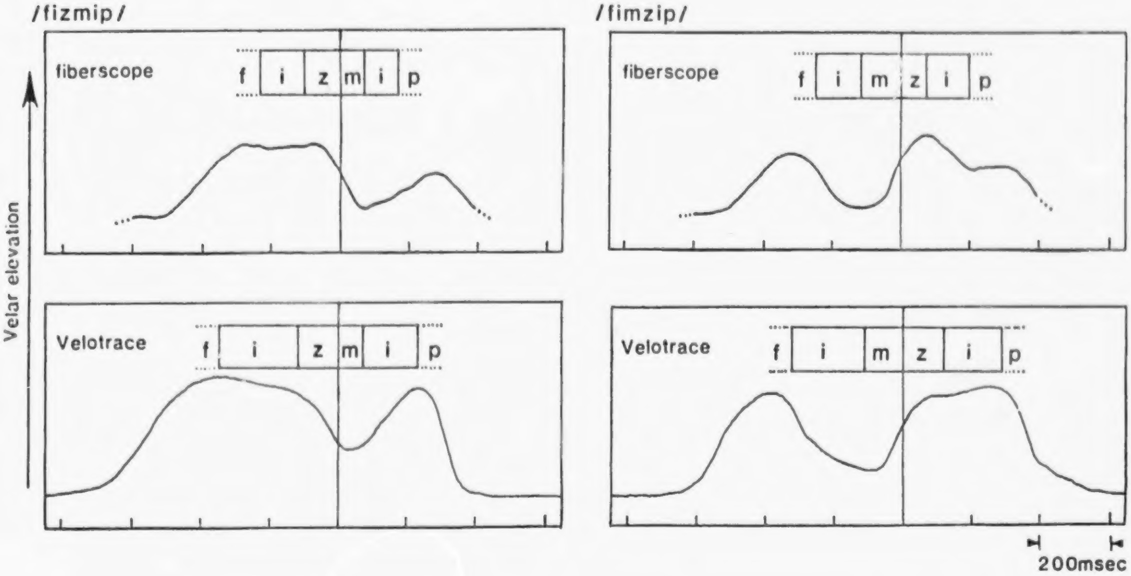


Figure 5. Ensemble-average endoscopic data from one subject (above) and Velotrace data from a second subject (below) for two utterance types (produced in isolation). Zero on the abscissa identifies the reference point for aligning tokens of an utterance type for computer sampling and averaging.

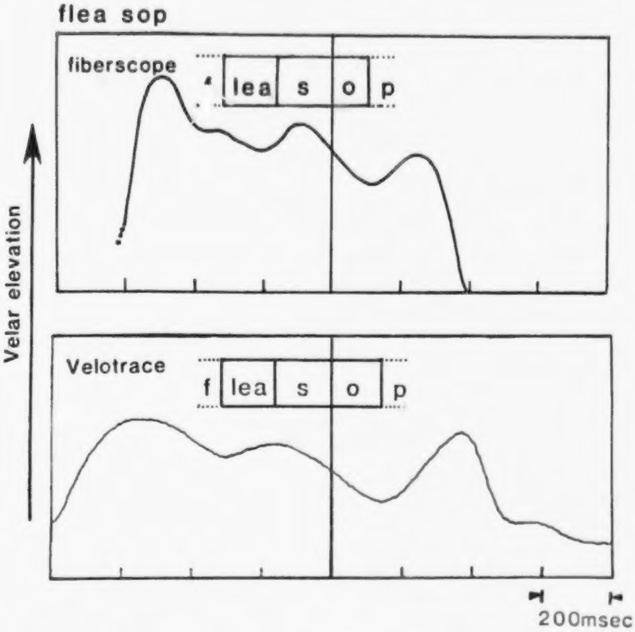


Figure 6. Ensemble-average endoscopic data from one subject (above) and Velotrace data from a second subject (below) for one utterance type (produced in a carrier phrase). Zero on the abscissa identifies the reference point for aligning tokens of an utterance type for computer sampling and averaging.

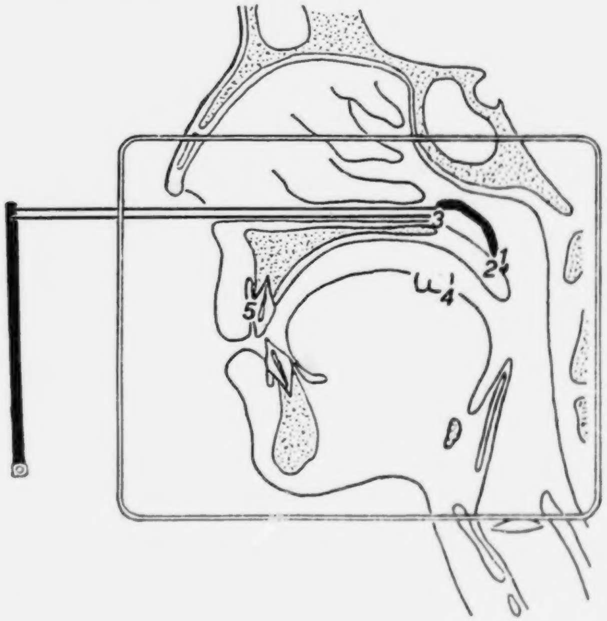


Figure 7. Schematic drawing of lateral cine X-ray film frame image with the Velotrace in place and measurement points indicated: (1) tip of the Velotrace internal lever; (2) the point on the velum that would be tracked by the Velotrace; (3) Velotrace internal lever fulcrum; (4) upper molar reference point; (5) upper incisor reference point.

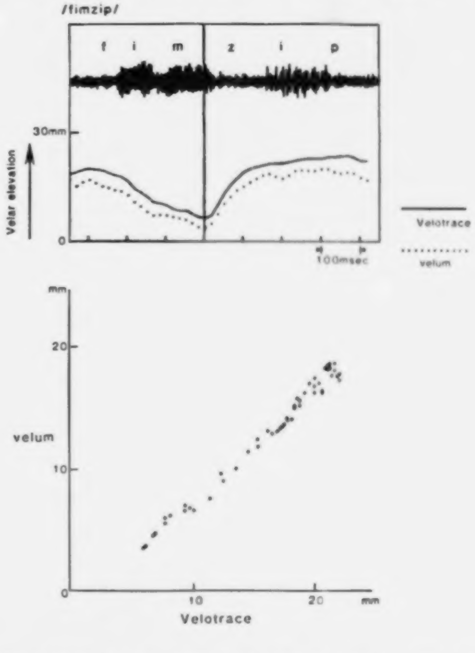


Figure 8. Comparison between the actual vertical movement of the velum and the movement of the tip of the internal lever of the Velotrace. The time function for one token is shown in the upper panel, with the acoustic waveform at the top and the Velotrace tip elevation (solid line) and velar elevation (dotted line) data below. A scatterplot of velar elevation and Velotrace tip elevation for each sample point in one token is shown in the lower panel.

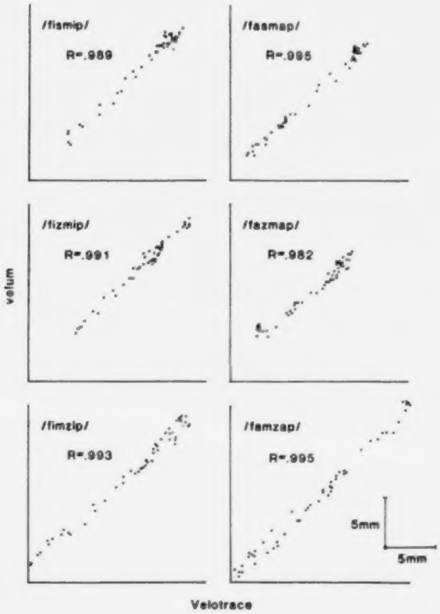


Figure 9. Scatterplots of vertical velar position vs. Velotrace position data for one token of each utterance type.

Conclusion

In evaluating the Velotrace, we compared data collected with the Velotrace with data obtained from frame-by-frame measurements of cine films (both endoscopic and radiographic) and found that the Velotrace signal accurately reflects the relatively rapid movements of the velum during speech. Among the advantages of the Velotrace as a device for monitoring the velum during speech are the elimination of X-ray exposure for the subject, and of frame-by-frame measurement for the investigator. Furthermore, the elimination of both of these drawbacks makes possible the collection and analysis of substantial quantities of data that should allow the development of a more complete understanding of velar motor control. In addition, the analog Velotrace signal can be sampled at a sufficiently high frequency to allow calculation of the highly accurate velocity and acceleration functions of velar movement patterns.

The Velotrace has a number of potential applications, among them the study of velar kinematics in normal speakers. Additionally, it may be used to monitor velar function in a number of different speech pathologies. For example, by carrying out clinical studies of individuals suffering from neuromuscular pathologies that affect velar function during speech and swallowing, one should be able to provide objective descriptions of the nature of the disruptions of speech and swallowing, although the Velotrace does not provide information about lateral pharyngeal wall movement. Such information should also provide further insight into the nature of the organizational patterns of velar motor control. Another application would be to the study of velar movement patterns in persons with velopharyngeal insufficiency, to examine the ways in which vertical movements of the velum differ from, or are similar to, those of normal speakers: That is, do they use "normal" or nearly normal articulatory strategies that fail because of anatomical and/or physiological limitations? Similar studies could be conducted with persons having mobile repaired clefts, to identify their articulatory strategies. Finally, the Velotrace may serve as a biofeedback device for training individuals with a variety of velar function problems, including pre-lingual hearing impairment, as well as the disorders mentioned above. We would note, however, that extending the use of the Velotrace to studies of children's speech depends upon considerations of instrumental and anatomical size, as well as interference of the adenoids with the function of the internal lever. Furthermore, the use of this device to study velopharyngeal function in persons with anatomical anomalies may require modification of the device.

References

Bell-Berti, F. (1976). An electromyographic study of velopharyngeal function in speech. Journal of Speech and Hearing Research, 19, 225-240.

Bell-Berti, F. (1980). A spatial-temporal model of velopharyngeal functions. In N. J. Lass (Ed.), Speech and language: Advances in basic research practice (Vol. IV). New York: Academic Press.

Bell-Berti, F., Baer, T., Harris, K. S., & Niimi, S. (1979). Coarticulatory effects of vowel quality on velar function. Phonetica, 36, 187-193.

Fritzell, B. (1969). The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. Acta Oto-laryngologica Suppl. 250.

Henderson, J. B. (1984) Velopharyngeal function in oral and nasal vowels: A cross-language study. Unpublished doctoral dissertation, University of Connecticut.

Kuehn, D. P. (1976). A cineradiographic investigation of velar movement variables in two normals. Cleft Palate Journal, 13, 88-103.

Moll, K. L., & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. Journal of the Acoustical Society of America, 50, 678-684.

Footnotes

[1] For Group 1 utterances, the beginning of [m] was chosen as the reference point (voicing onset following [p]; voicing onset or end of frication following [s]; increased amplitude following [b]; end of frication following [z]). For Group 2 utterances, the end of [m] was chosen as the acoustic reference point (voicing offset before [p]; voicing offset or beginning of frication before [s]; amplitude reduction or voicing offset before [b]; beginning of frication before [z]). For Group 3 utterances, the end of the medial consonant-sequence occlusion was chosen as the acoustic reference point (end of frication for [...sV] and the stop burst for [...tV] utterances).

[2] As a result of field-size limitations and because we were primarily interested in knowing how well the internal lever follows movements of the velum (rather than how well the external lever reflects movements of the internal lever), the external lever was not included in the viewing field.

[3] Kuehn's displacement-versus-time data, taken from the constant velocity portion of displacement-versus-time curves, are not directly comparable with our data for two reasons. First, his data are measures of Euclidean

distance, whereas ours are of vertical distance only. Second, his data are an index of the velocity during the relatively constant velocity portion of the gesture, whereas ours are the peak values in the first derivatives of our displacement-versus-time functions. However, using the angular factor that he reported, we have estimated the maximum y-trajectory velocities for each of his two subjects. The maximum y-trajectory upward velocities are 54mm/sec and 90mm/sec; downward velocities are 38mm/sec and 83mm/sec.

TOWARDS AN ARTICULATORY PHONOLOGY*

Catherine P. Browman and Louis M. Goldstein†

Abstract. We propose an approach to phonological representation based on describing an utterance as an organized pattern of overlapping articulatory gestures. Because movement is inherent in our definition of gestures, these gestural "constellations" can account for both spatial and temporal properties of speech in a relatively simple way. At the same time, taken as phonological representations, such gestural analyses offer many of the same advantages provided by recent nonlinear phonological theories, and we give examples of how gestural analyses simplify the description of such "complex segments" as /s/-stop clusters and prenasalized stops. Thus, gestural structures can be seen as providing a principled link between phonological and physical description.

1. Introduction

The gap between the linguistic and physical structure of speech has always been difficult for phonological theory to bridge. Until recently, theories have encapsulated the linguistically-relevant structure of speech in a sequence of segmental units, each of which corresponds to a feature bundle. Under this strict segmental hypothesis (formulated in terms of features), the sequence of feature bundles that constitute segments forms a feature matrix, whose cells are organized into non-overlapping columns. Linguistically relevant contrast between utterances, in this approach, requires that at least one feature value differ between contrasting strings. The bridge to the continuous nature of speech is made by assuming that "each segment is characterized in terms of a state of the vocal organs, and the transitions between these states are ... predictable in terms of very general linguistic and physiological laws" (Anderson, 1974, p. 5).

This strictly linear view of the relation between linguistic units and speech has come under attack in recent years from two different directions. Phonologists have found the constraint imposed by linear sequences of non-overlapping segments to be too extreme to capture a variety of phonological facts. Recognition of the importance of allowing feature specifications to overlap was made, e.g., by Anderson (1974). He presented an alternative approach that decomposed articulation into four subsystems (an

---

*In press, Phonology Yearbook (Vol. 3, 1986).

†Also Yale University

Acknowledgment. This paper has benefitted greatly from the comments and criticisms of several dozen colleagues, primarily from Haskins, Yale, and UCLA. We only wish we could thank each of them individually here. Any weaknesses remaining in the paper are due solely to our own intransigence in the face of their patient and generous critiques. This work was supported in part by NIH grants HD-01994, NS-13870, and NS-13617 to Haskins Laboratories.

energy source, a laryngeal system, an oral system, and a nasal system), and noted that "it is possible... [that] the boundaries of specification in one system will not coincide with the boundaries of a specification in another" (1974, p. 274). The other direction of attack has come from phoneticians (e.g., Lisker, 1974), who have shown the linguistic relevance of the detailed temporal structure of speech. For example, as discussed below, interarticulator temporal organization may vary from language to language in a way that cannot be predicted (by any universal principles) from existing phonetic feature characterizations, and thus, must be specified somehow in language descriptions. These developments suggest a need for a revised conception of phonological/phonetic structure, one that incorporates overlapping phonological units and one that allows temporal relations among articulatory structures to emerge from the description. We consider the two lines of attack in greater detail.

The linearity assumption has been challenged (if not completely discarded) by attempts over the last ten years to formalize more enriched conceptions of phonological structure. Like Anderson's (1974) proposal, these efforts were undertaken in response to the failure of the segmental model to account adequately for certain facts. These conceptions include explicit incorporation of syllable structure (Hooper, 1972, 1976; Kahn, 1976), hierarchical metrical structures (Hayes, 1981; Liberman & Prince, 1977; Selkirk, 1980), dependency structures (Anderson & Jones, 1974; Ewen, 1982), independent structural or autosegmental tiers (Clements, 1980; Goldsmith, 1976), and explicit incorporation of a consonant-vowel skeleton (Clements & Keyser, 1983; Halle & Vergnaud, 1980; McCarthy, 1981, 1984; Prince, 1984). While these approaches have increased the range of facts that can be adequately formalized in phonological theory, they are inexplicit with respect to the relation between the revised conception of phonological structure and the physical structure of speech. The traditional link between phonological and physical structure has vanished along with strictly linear segmental analyses, and a new link has yet to be forged. It is this task we attempt in this paper, by accounting as simply as possible for the organization of speech in both space and time. We will show that the structures that emerge from such an account can also be used as a basis for phonological description--indeed, a kind of phonological description that is much in the spirit of the above-mentioned theories.

From the phonetic side, there has been growing evidence that systematic phonetic feature representations cannot adequately describe phonetic differences among languages. Ladefoged (1980), for example, argues that the specification of features at the systematic phonetic level is neither "necessary nor sufficient to specify what it is that makes English sound like English rather than German" (1980, p. 495). As both Ladefoged and Anderson (1974) point out, phonetic differences between languages may involve aspects of speech that do not serve as the basis for phonological contrast within any one language. For example, Anderson shows that languages differ with respect to whether stops are released in clusters and in word final position, even though no single language contrasts released vs. unreleased stops. Thus, he suggests a feature [±release] to differentiate the phonetic representations in these languages.

The difference between released and unreleased stops can be seen as part of a more general problem: differences among languages in the relative timing of articulatory gestures. The "unreleased" initial stops in clusters are, presumably, released, but only after the occlusion for the second stop has formed. There would be little acoustic evidence, therefore, of their release (cf. Catford, 1977). Thus, language differences in stop release may be analyzed as differences in the temporal overlap of adjacent closure gestures. It is possible, in general, to describe such cross-language differences in gestural timing within the SPE framework (Chomsky & Halle, 1968) by means of features such as [±release]. However, the potential number and variety of such differences would lead to the proliferation of features that have no contrastive function within languages. (A similiar point about proliferation of phonetic features, but not specifically about timing relations, is made by Keating, 1984).

It is not difficult to find documented examples of cross-language differences in gestural timing. A number of writers (e.g., Flege & Port, 1981; Keating, 1985; Mitleb, 1984; Port, 1981) have demonstrated such differences in voicing contrasts, specifically in the duration of vowels preceding voiced and voiceless stops. While the acoustic duration of a preconsonantal vowel is generally longer before a voiced than before a voiceless stop, the effect is larger in some languages than in others (as was earlier noted by Lehiste, 1970), and can be virtually absent (e.g., in Polish, Czech, and Arabic). These differences in vowel duration presumably reflect differences in the relative timing of vowel and consonant gestures. In this case, a different feature, probably [±long], would be used in an SPE treatment to describe cross-language differences in gestural timing.

Such phenomena are not restricted to voicing. For example, languages may also show differences in ejective consonants in the time between release of glottal closure and release of oral closure (Catford, 1977, p. 69; Lindau, 1984). There is no evidence that such differences in ejectives contrast in any language, and therefore, some ad hoc feature, similar to [±release], would have to be proposed to account for this difference. Finally, Fourakis (1980) has shown that the occurrence of so-called epenthetic stops in English words like <tense> is dialect-dependent. If such "stops" are to be analyzed in terms of variation in the relative timing of oral and velic gestures (rather than actual segment insertion, cf. Anderson, 1976; Ohala, 1974), then such timing relations are also not universal, but are a property of a particular language or dialect.

In general, then, languages can differ from one another in the timing of (roughly the same) articulatory gestures. The above examples are meant simply to illustrate the variety of phenomena that can be analyzed in this way. An SPE characterization of such examples not only proliferates features in the grammar; more seriously, it misses a generalization: timing of articulatory gestures is linguistically relevant, at least in terms of how languages are distinguished from one another.

One approach to the description of language-particular patterns of articulatory timing has involved positing rules that specifically convert segmental feature matrices to temporally continuous physical parameters (e.g., Keating, 1985; Port, 1981). As described by Port and O'Dell (1984), such "implementation" rules are like phonological rules in that they must be assumed to differ from language to language, but unlike such rules, they do not map feature values onto feature values. Rather they take features (or matrices thereof) as input and output "a complex pattern of graded commands distributed over time to the various articulators" (Port & O'Dell, 1984, p. 122). However, such rules for implementing patterns of articulatory timing have not been made explicit. This is not surprising, perhaps, since the supposed output of the rules--control parameters for articulators--requires reference to the organization of articulatory movements. Yet no linguistic approach has provided a vocabulary for describing such organization. In the implementation rule view, the organization remains outside of speech itself, in the segmental (and metrical, etc.) structure. Speech itself has no organization, but is rather seen as a plastic medium that somehow serves to code the information present externally in the linguistic structure. It is worth noting that the implementation rules that have been successfully made explicit by Liberman and Pierrehumbert (1984) characterize intonation. Here, the relevant physical parameter is univariate in the acoustic domain (F_0), and no assumptions about articulatory organization are made.

Rather than positing inexplicit implementation rules or proliferating ad hoc features, we propose to base phonological representation on an explicit and direct description of articulatory movement in space and over time. As argued by Fowler (1980), incorporation of time (in particular) into the basic definition of phonetic units can simplify much of the complex translation that is required in an approach like that of implementation rules. Moreover, describing speech in terms of overlapping (relatively invariant) articulatory units with inherent time courses can account in a simple way for observed patterns of acoustic variation (Bell-Berti & Harris, 1981; Fowler, 1980; Fujimura, 1981; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Mattingly, 1981). Such articulatory descriptions are particularly promising for nonlinear phonological analyses that require overlapping features, because there is a clear physical reality underlying the decomposition of articulation into quasi-independent systems whose movements are not always synchronized.

While there is, of course, a long history of referring to aspects of articulation in phonological and phonetic representations (e.g., Abercrombie, 1967; Chomsky & Halle, 1968; Jespersen, 1914; Ladefoged, 1971; Pike, 1943), such representations have often been forced to rely on impressionistic descriptions, and have emphasized the static aspects of articulation. Two recent developments in speech research make it feasible, we believe, to incorporate explicit characterizations of articulatory movement into phonological representation. The first of these is the development of improved technologies for tracking continuous articulatory movement (e.g., Fujimura, Kiritani, & Ishida, 1973), which reduces the need to rely on impressionistic observations by providing more explicit physical measurements. The second, the development of a theoretical framework for the analytical and mathematical description of coordinated movements (e.g., Bernstein, 1967; Fowler et al., 1980; Kelso & Tuller, 1984a, 1984b; Kugler, Kelso, & Turvey, 1980; Saltzman & Kelso, 1983; Turvey, 1977), simplifies the description of movement sufficiently to make it tractable for phonological purposes, and also provides a framework in which to explore generalizations regarding coordinated articulatory movements.

As we will attempt to show in this paper, an explicit description of articulatory movement can serve as the basis for phonological representation. The basic units in this framework are articulatory gestures (section 1.1). We will first define gestures simply as characteristic patterns of movement of vocal-tract articulators, or articulatory systems, and will suggest phonological analyses of two linguistic problems in terms of these gestures and the relations among them (section 2). Such analyses have many of the advantages found in recent nonlinear phonological theories, while at the same time providing a possible solution to the problem of the missing link between phonological and physical structures. We will then show how these characteristic patterns of movement can emerge from an abstract mathematical formalization of gestures and the lexical structures composed of such gestures (section 3). This mathematical formalization of a gesture, being developed in cooperation with our colleagues at Haskins Laboratories, uses a dynamical model to explicitly characterize the coordinated patterns of articulatory movement. In this approach, based on the concept of coordinative structures (e.g., Turvey, 1977) as instantiated in the task dynamic model of Saltzman & Kelso (1983), gestures are autonomous structures that can generate articulatory trajectories in space and time without any additional interpretation or implementation rules.

1.1 Gestural Structure of Speech

We begin our discussion of gestures with a simple example, the utterance [abə]. During this utterance, the lower lip moves up gradually toward the upper lip, reaches some peak upward displacement, and then moves downward again, as can be seen in Figure 1. The lip is constantly in motion, except instantaneously at its maximum displacement--it does not necessarily achieve any steady-state configuration that could be associated unambiguously with the /b/. The absence of steady states characterizes all observations of speech, whether articulatory or acoustic, and this may be one of the reasons why phonologists have posited a complex relation between phonology and speech. We argue, however, that it is the assumption of a steady state specification that leads to the apparent complexity, and that a phonology which inherently incorporates movement in its descriptions will simplify much of this apparent complexity.

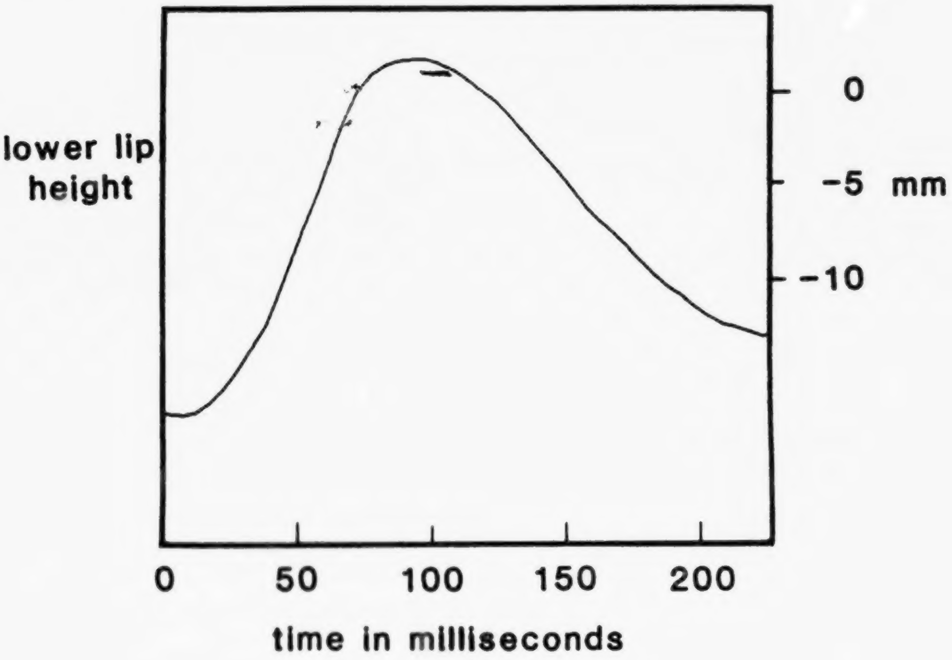


Figure 1. Trajectory of the lower lip in [abə], as measured by tracking infrared LED placed on subject's lower lip.

Instead of looking for a steady-state correlate of the /b/ segment, then, we take the trajectory of the lower lip in [abə] as a pattern in space and time that characterizes utterances transcribed with a /b/ in them. That is, there is no particular spatial coordinate value of the lower lip (or any other articulator) that is held for some time and that is characteristic of /b/. Rather, it is the movement of the articulators through space over time that constitutes an organized, repeatable, linguistically relevant pattern. We can refer to this pattern as a bilabial closure gesture. [Not every utterance of a word transcribed with a /b/ will display exactly the trajectory of Figure 1: the trajectory will vary with vowel context, syllable position, stress, speaking rate, and speaker. We must, therefore, ultimately characterize a /b/ as a family of patterns of lip movement. In section 3, we will suggest how this family can be formally defined using an abstract set of equations that can generate the variant trajectories. For the present, we can think of a gesture as an instance of a family of related trajectories.]

If every segment in a traditional phonological (or phonetic) representation could be described as one gesture, much as /b/ can be described as a bilabial closure gesture, then the implications of the gestural approach for phonology would be limited. However, the relation between segments and gestures is not always one-to-one. English voiced stops can, to a first approximation, be characterized as single gestures of bilabial, alveolar, or velar closure; but other segments, such as the voiceless stops, require more than one gesture. In /p/, for example, we have a bilabial closure gesture much like that for /b/. In addition, however, the glottis must be opened for /p/, and then narrowed again. That is, from the point of view of spatio-temporal speech structure, /p/ is an organization of two gestures--a bilabial closure gesture plus a glottal opening (and closing) gesture. Thus, there is no one-to-one relation between gestures and segments.

Nor do gestures bear a one-to-one relation to traditional phonological features. A single bilabial closure gesture would correspond to a number of features, such as [-continuant], [+anterior], [-coronal], [+consonantal], [-vocalic], etc. In general, differences in the presence or absence of glottal or velic (opening and closing) gestures correspond to single feature differences, while supraglottal constriction gestures correspond to multiple feature differences. Thus, gestures do not bear a one-to-one relation to either phonological segments or phonological features. Rather, they represent organized patterns of movement within oral, laryngeal, and nasal articulatory systems.[1]

In addition to the gestures themselves, the relations among gestures also play an important role in the articulatory description, similar to the role of the associations among autosegments in autosegmental phonology. As an example, consider the bilabial closure and glottal opening gestures in words transcribed as beginning with /p/. These gestures are not temporally simultaneous, but repeated observations of words beginning with /p/ reveal tight spatial and temporal relations between the two gestures (Löfqvist, 1980; Löfqvist & Yoshioka, 1985; cf. Lisker & Abramson, 1964). The incorporation of such spatio-temporal coordination within our description can be seen as having two different functions. On the one hand, the representation of a tight relation between the two gestures defines a phonological class--the class traditionally described as words beginning with /p/. On the other hand, the relation is specified in explicit enough fashion to capture the systematic, language-particular aspects of the timing between the gestures.

In the particular example of /p/, the gestural structure specified corresponds to a segment. In general, however, the interdependencies among gestures are not restricted to those that constitute single segments in traditional approaches. Rather, the pattern of relations among a set of gestures, the gestural constellation, can serve the functions typically filled by other phonological structures, ranging from complex segments to syllables and their constituents. In section 2, we show how such constellations can provide the basis for phonological analyses in cases where featural overlap has been invoked, of the type explored by Anderson (1974) and developed further within autosegmental, including CV and X-tier, phonologies.

2. Gestural Analysis of Two Linguistic Problems

As noted in the Introduction, segmental and gestural analyses differ minimally for gesturally simple segments like /b,d,g/. Even in such cases, however, a gestural analysis has the additional advantage that it accounts for the physical movements of the articulators as well as for the phonological structure. For gesturally more complex structures, the gestural analysis differs from a segmental analysis. In this section, we present two instances in which the analyses diverge: English /s/-stop clusters (section 2.1) and prenasalized stops (section 2.2). Both are examples of "complex segments" (e.g., Ewen, 1982), that is, they behave in some ways like single segments and in some ways like clusters. Using the gestural approach, we attempt to answer the question of "one" vs. "two" units by analyzing the observed articulatory movements themselves. As we shall see, this gestural analysis can provide structures that allow linguistic facts to be stated more generally and simply than in a segmental analysis. We will also see how these same gestural structures can account for some of the observed patterns of timing.

2.1 Glottal Gestures and /s/-Stop Clusters

In a segmental phonology, the description of initial /s/-stop clusters in English is problematic. There are at least two facts about these clusters that require specific statements within the phonology--statements that apply only to these clusters. First, the phonotactics must state that there is no contrast between voiced and voiceless stops following initial /s/; that is, there is a "defective" distribution. Second, the realization of this stop as voiceless unaspirated must be specified by a separate phonetic (or phonological) rule. In current approaches, these facts do not follow from any more general characterizations of English phonology or phonetics.[2] In addition, other problematic aspects of such clusters have led phonologists to argue that they are more "unitary," i.e., more nearly describable as single segments, than are other clusters. [Ewen (1982) summarizes the evidence for a monosegmental analysis, which includes the /s/-stop clusters' violation of the sonority hierarchy and the failure of these clusters to alliterate with /s/ in Germanic verse.] In this section, we will show not only how, in an articulatory phonology, the phonetic and distributional facts about the clusters follow from a more general constraint on the articulatory structure of English words, but also how the gestural structure might account for the clusters' ambiguous status as one or two units.

Crucial to this account is an understanding of the behavior of the glottis in voiceless stops and clusters. This has recently been investigated in English (Yoshioka, Löfqvist, & Hirose, 1981), Swedish (Löfqvist & Yoshioka, 1980a), Icelandic (Löfqvist & Yoshioka, 1980b; Petursson, 1977), and Danish (Fukui & Hirose, 1983). Like English, these other Germanic languages contrast an initial voiceless aspirated stop with either an unaspirated or voiced initial stop, but neutralise the contrast after /s/. In all cases, a single glottal opening/closing gesture is found for words beginning with /sC/ clusters (where C is a stop). This single gesture is similar to the one that occurs either with /s/ alone initially, or with one of the initial voiceless aspirated stops, although the magnitude of the gesture tends to be a bit smaller when accompanying the stops. As Petursson (1977) argues, the failure to find two glottal gestures in the initial /sC/ clusters cannot be due to a principle of economy of movement under which the glottis remains open for as many voiceless segments as are required. That is, the glottis is not, apparently, held in an open position during long periods of voicelessness (Yoshioka et al., 1981), but rather exhibits a sequence of opening and closing movements. This can be found, for example, in /s#C/ sequences (i.e., sequences containing a word boundary), which can show two glottal opening and closing gestures. Thus, the failure to find two glottal gestures for initial /sC/ clusters points to a generalization about the linguistic organization in these languages--words begin with, at most, a single glottal gesture.

These observations form the basis for the gestural analysis of /sC/ clusters. Here, contrasts are described in terms of different characteristic gesture constellations--that is, one or more gestures in a specific spatio-temporal relation. For example, /pa/ and /ba/ differ in the presence vs. absence of the glottal opening/closing gesture. /pa/ and /sa/ differ in that one has a bilabial closure gesture, the other an alveolar fricative gesture, both in constellations with a glottal gesture (with characteristic spatio-temporal relations). The initial /sp/ cluster is a constellation of three gestures--an alveolar fricative gesture, a bilabial closure gesture and a single glottal opening/closing gesture. Thus, the glottal gesture can occur in a constellation with a single oral constriction gesture (as in /pa/ or /sa/), with two (as in /spa/), or alone (as in /ha/). [cf. Hockett's (1955) proposal of an immediate constituent analysis for /sC/ clusters along similar lines.] The phonotactics of English in this approach is a statement of the possible constellations of gestures. The generalization of interest here is that, in word-initial position, English has at most one glottal gesture, of roughly constant magnitude, regardless of the other gestures with which it co-occurs.

This generalization accounts for the lack of word-initial contrast in English between /sp/ and /sb/. That is, a word-initial contrast would require either two glottal gestures in the constellation for /sp/, or a different, much smaller, glottal gesture for /sb/. Both of these possibilities are ruled out by the single-glottal-gesture generalization. Moreover, in conjunction with a fact about intergestural coordination in English, it also accounts for the realization of the /p/ in initial /sC/ clusters as voiceless unaspirated. This additional fact is that peak glottal opening typically occurs at the midpoint of a fricative gesture, if there is one present in the constellation, or, if not, at the release of a stop closure gesture (Yoshioka, Löfqvist, & Hirose, 1981). The generalization is presented in its current form in (1); it might ultimately be simplified by referring to the coordination of the glottal gesture with the vowel.

(1) Glottal gesture coordination in English

(a) If a fricative gesture is present, coordinate the peak glottal opening with the midpoint of the fricative.

(b) Otherwise, coordinate peak glottal opening with the release of the stop gesture.

Statement (1) holds for both single consonants and consonant clusters. For single consonants, it accounts for the fact that initial voiceless stops are aspirated (since, by (1b), the peak glottal opening occurs at the release of closure), while initial fricatives are not (1a). For clusters like /sp/, it accounts for the lack of aspiration. That is, since the peak glottal opening occurs during the fricative (1a), by the time the following stop closure is released the glottis is already narrowed, producing a voiceless unaspirated stop. [The above analysis of aspiration is similar to a proposal by Catford (1977), who did not, however, explicitly discuss gestural organization.]

Statement (1) can also account for another aspect of the phonetic structure of English--the devoicing of sonorants following initial voiceless stops but not following initial voiceless fricatives. For initial /pl/, for example, (1b) predicts that the peak of the glottal gesture is timed to occur at the release of the stop gesture, regardless of the presence or absence of other gestures. As Catford (1977) notes, the alveolar lateral (impressionistically, at least) has already been achieved by the time of release of the stop closure, so that the wide-open glottis co-occurs with the lateral, producing a voiceless lateral. Thus, the voicelessness of the lateral follows directly from the nature of the gestures and the independently required generalization about gestural coordination, and does not have to be stated as a separate allophonic rule. In contrast, (1a) predicts that sonorants will be only slightly devoiced following initial voiceless fricatives, and not devoiced at all in clusters such as /spl/, since the peak glottal opening occurs at the midpoint of a fricative gesture regardless of the number of following consonantal gestures. Thus, the intergestural coordination generalization captures a number of facts about English phonetics that would otherwise require separate statements.

Returning to the single-glottal-gesture generalization, we note that it also has implications beyond its explanation of the defective distribution of /sC/ clusters. The ambiguous nature of such clusters is inherent in their proposed gestural constellations, consisting of a single glottal gesture with two overlapping oral gestures. These clusters might act as single units under the influence of the single glottal gesture, or as sequences of two units under the influence of the two oral gestures. A similar analysis has been proposed for ejective clusters such as [t'p'] in Kabardian (Anderson, 1978). He argues that Kuipers' (1976) analysis of Kabardian as phonologically vowelless can be maintained if the ejective clusters are treated in a unitary way as complex segments. Their unitary phonological behavior is related by Anderson to their articulatory nature, consisting of a sequence of two oral articulations associated with a single laryngeal gesture. Thus, he proposes an autosegmental-type analysis, in which a single laryngeal specification is associated with a sequence of specifications for oral articulators. The parallel with the gestural constellations proposed here for /sC/ clusters is obvious. Both cases independently support our contention that gestural structures derived from observing articulatory movements provide an appropriate basis for stating phonological generalizations. Moreover, taken together, they suggest a general principle that a particular type of gestural structure (one laryngeal gesture organized with two oral ones) may be associated with ambiguous phonological behavior.

2.2 Prenasalized Stops and Nasal-Stop Clusters

Prenasalized stops constitute another class of complex segments whose analysis has been used to enrich the strictly segmental phonological model (e.g., Anderson, 1976; Ewen, 1982; Feinstein, 1979). In this section, we will show that the difference between prenasalized stops and nasal-stop sequences posited in such analyses cannot predict the kinds of temporal regularities shown by nasal-stop sequences in English, unless certain *ad hoc* rules are added. These temporal regularities lead us to hypothesize a *similar* gestural analysis for prenasalized stops and English clusters, and we will present articulatory data to support this analysis. This gestural analysis, we will argue, captures the relevant phonological generalizations while allowing the temporal regularities to be predicted directly from the gestural organization.

Anderson (1976) has presented arguments for analyzing prenasalized stops as single segments, but with a sequence of values for the feature [nasal] (this is consistent with Herbert's 1975 acoustic data showing that prenasalized stops have roughly the same duration as simple stops). Thus, the domain of value-assignment for the nasal feature is not coterminous with the boundaries between segments. In this way, the ambiguous nature (unitary vs. sequential) of such stops can be directly captured. His representation for a prenasalized stop is shown in (2a) and for a sequence of homorganic nasal + stop in (2b).

(2)	(a) m b	(b) m b
cons	+	+ +
nasal	+ -	+ -
ant	+	+ +
cor	-	- -
.		
.		

The structures represented in (2a) and (2b) might be expected to lead to different phonetic entities. From the gestural point of view, we would expect to find a difference between the bilabial closure gestures in (2a) and (2b), with the prenasalized stop (2a) having a single bilabial closure gesture, and the nasal-stop cluster (2b) having either two bilabial closure gestures, or, possibly, a single longer bilabial closure gesture. Since in English, words like <camper> and <canker> are analyzed as having nasal-stop sequences, we would expect them to have structures like that in (2b). [The analysis is supported by distributional considerations: /mp/, /mb/, etc. cannot occur in syllable-initial position where no sonorant-stop sequences are allowed, but they can occur post-vocalically, along with other sonorant-stop sequences.] However, certain durational properties of the words containing these clusters, specifically the duration of the preceding vowels, are not correctly predicted by representations such as (2b).

The durational characteristics of the English words <camper> and <camber> were analyzed acoustically by Vatikiotis-Bateson (1984) and compared to words containing single segments, <capper>, <cabber>, and <cammer>. His results showed, as expected from other studies (e.g., Haggard, 1973; Lindblom & Rapp, 1973; Walsh & Parker, 1982), shortening of the nasal and stop segments when they occurred in a cluster, compared to their durations as single consonants. However, contrary to expectations based on other studies (e.g., Fowler, 1983; Lindblom & Rapp, 1973), the stressed vowels preceding the nasal-stop clusters did not shorten when compared to the vowels before single consonants (i.e.,

/mp/ vs. /p/ and /mb/ vs. /b/). That is, the labial nasal-stop sequences behaved like single consonants in terms of their effects on preceding vowel duration. Moreover, this similarity of behavior between the clusters and the single consonants extended to the effect on the preceding vowel duration of the consonantal voicing. In the Vatikiotis-Bateson data (as has also been shown by Lovins, 1978, and Raphael, Dorman, Freeman, & Tobin, 1975), vowels were shorter before the clusters containing a voiceless stop as well as before the single voiceless consonants, in spite of the fact that the nasal immediately following the vowel was voiced. That is, it was the voicing of the oral stop that determined the vowel length differences.

In the above analyses of acoustic duration, the movement of the velum appeared to be irrelevant, since nasal-stop clusters behaved like single consonants. The similarity between the effects of clusters and singletons could be accounted for if there were a single bilabial gesture in English bilabials and nasal-stop sequences, regardless of the movement of the velum. Such a specification is best captured by representation (2a). This implies in turn, however, that nasal-stop sequences in English should have the same gestural structure as prenasalized stops.

To test this hypothesis about the gestural structure of English nasal-stop sequences, we collected articulatory data for the same set of words containing bilabial stops and nasal-stop sequences that Vatikiotis-Bateson [illegible] speaker of Chaga. The same [illegible] was used for both speakers. Each spoke the selected words containing bilabial sequences in a carrier phrase, repeating the words in the phrases five times. The selected words and carrier phrases are listed in (3). Notice that there is no /baka/, since all voiced obstruents in Chaga are prenasalized.

(3)		*English*	*Chaga*
	phrase:	*It's a __ tomorrow.*	*/wia mboka __ kimbuho/* *'say to the starter __ slowly'*
		capper	/paka/ 'cat'
		cammer	/maka/ 'year'
	words:	cabber	
		camper	/mpaka/ 'boundary'
		camber	/mbaka/ 'curse'

The articulatory information for both speakers was gathered with a Selspot system at Haskins Laboratories. To track the movement of the lips and jaw, miniature infra-red light-emitting diodes (LEDs) were attached to the midpoint of the upper lip, the lower lip, and just under the chin (or, in the case of the bearded Chaga speaker, slightly above the chin). A modified video camera positioned to capture the profile of the speaker then tracked the movement of the diodes. In addition to the lip and jaw movement data gathered by the Selspot equipment, a gross measure of nasal airflow (during voiced speech) was obtained using an accelerometer attached to the bridge of the nose. In

Figure 2 shows the data for the English word <camper> in a carrier phrase. The acoustic signal is displayed in the bottom panel. The markings in each panel are derived from the acoustic signal--for example, NAS to indicate the onset of nasal murmur associated with the /m/, CLO for the onset of silence during the stop gap of the /p/, RL for the acoustic release of closure, AE and ER for voiced vocalic onsets. The top panel shows the information about nasality gathered by the accelerometer. Notice that, in addition to the nasal murmur (from 390 ms to 440 ms), the entire vowel /æ/ (from 290 to 390 ms) is nasalized. The middle panels display information about the vertical position of the upper and lower lips over time. Here we are most concerned with the labial closure gesture, extending approximately from 350 to 500 ms. The upper lip has very little vertical movement, while the lower lip smoothly raises and then lowers for closure and release. The actual acoustic closure encompasses most of the peak of the labial gesture, from 390 ms (beginning of the nasal murmur) to 470 ms (release of the stop). [Note that closure is achieved before the highest point in the articulatory movement is achieved, and is not released (point labelled RL) until after the lip begins to move down again. The lips are, therefore, continuing to move even during acoustic closure, presumably compressing the tissues involved.] We will be concentrating on this lower lip movement as we investigate the articulation of the various labial sequences.

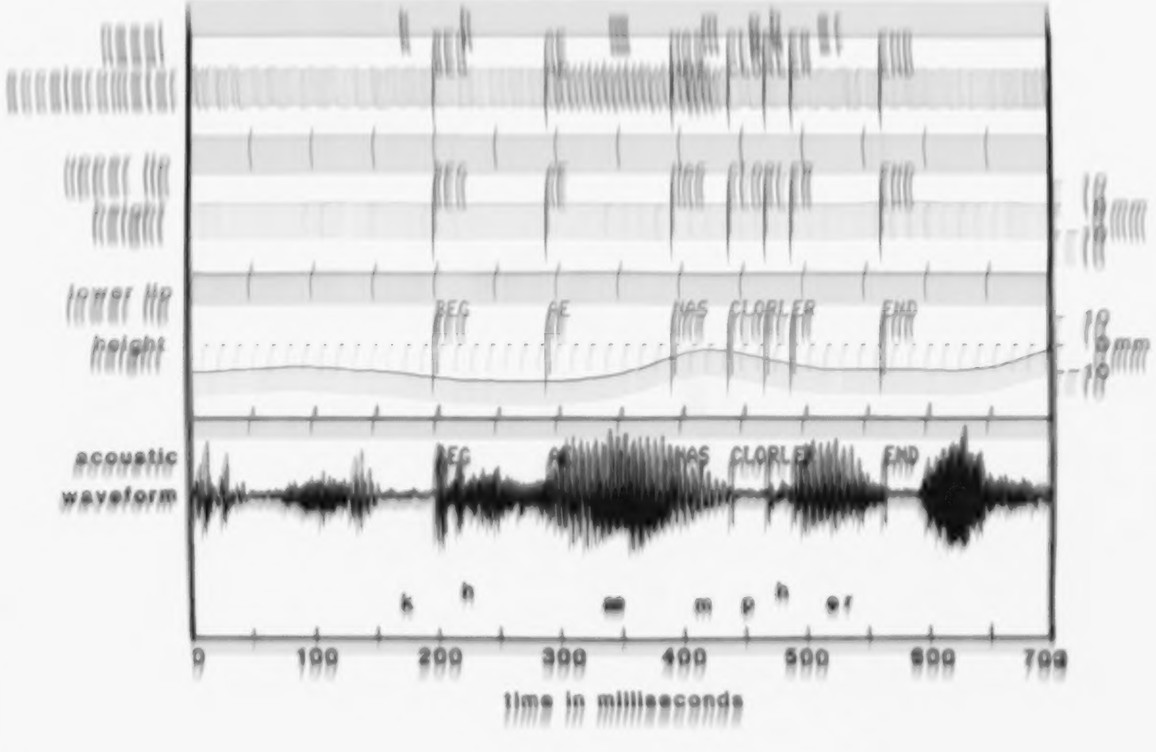


Figure 2. Acoustic waveform and articulatory measurements for single token of <camper>.

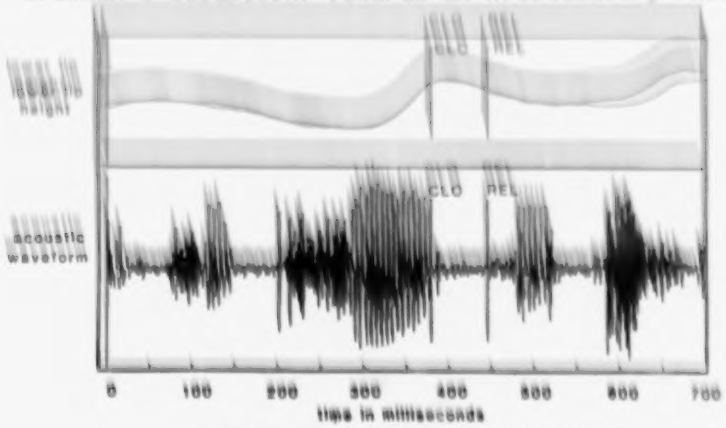
Our primary interest in the labial gestures is in the similarities and differences among gestures in different phonological categories, that is, single consonants, prenasalized consonants, nasal-stop clusters, and syllabic nasal plus stop. Both speakers proved to be quite regular across tokens within the same phonological category. Typical examples are shown in Figure 3a for the English speaker (the medial /p/ of <capper>), and Figure 4a for the Chaga speaker (the initial /m/ of /maka/). In both figures, lower lip gestures for two repetitions of the word are superimposed and displayed above the acoustic signal from one of the tokens. The vertical lines indicate the onset and release of closure, as measured in the acoustic waveform. To compare tokens between phonological categories, a single repetition was chosen from each set of five for each of the words to be compared. In each case, this representative item (selected from the second, third, or fourth repetition) was identical, as determined by visual inspection, to at least one other repetition, both in terms of the pattern over time of the lower lip gesture, and in terms of the timing of the gesture relative to the surrounding vowels. These representative items are used in the rest of the figures; the conclusions based on these items have been confirmed by comparisons among all the repetitions.

The between-category comparisons indicate that, contrary to expectations based on segmental descriptions such as (2b), all of the phonological categories except for the syllabic nasal, are represented by a single labial gesture. That is, there is no observable difference among the labial gestures associated with a single consonant and a consonant cluster. This can be seen in Figure 3b for English, and Figure 4b for Chaga.

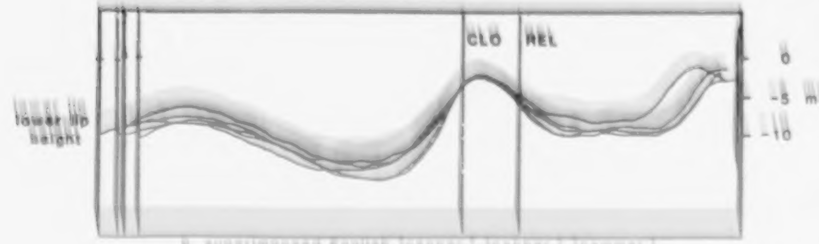
Figure 3b shows the lower lip traces for English <capper>, <cammer>, <camber>, and <camper>, superimposed on the trace for <capper>. (The gestures have been slightly offset, both horizontally and vertically, to facilitate comparison of their overall forms; the extent of the horizontal offsets can be determined from the lines on the left; the vertical offsets are represented by the tick marks on these lines.) While there are small differences in the amplitude and in the slope of the onset and offset of the gesture, which correspond to similar differences among /p/, /b/, and /m/ reported in the literature (e.g., Kent & Moll, 1969; Sussman, MacNeilage, & Hanson, 1973), the overall envelope of the gestures is similar, particularly in the central portion demarcated by the lines on the <capper> trace. That is, regardless of whether the consonantal portion is described as a single consonant (/b/, /p/, or /m/) or as a consonant cluster (/mp/ or /mb/), in English there appears to be a single labial gesture.

Figure 4b shows the superimposed lower lip gestures for the Chaga words /paka/, /maka/, and /mbaka/. Again, as in English, there is a single gesture, quite similar in overall envelope, and particularly in the central portion. That is, in Chaga there is a single labial gesture associated with single and prenasalized consonants.

The syllabic nasal+stop /mp/ in Chaga, however, presents a different picture. Comparing /maka/ and /mpaka/ in Figure 5a, we see for the first time a clear difference in the overall duration of the envelope of the lower lip gesture. The gesture for /mpaka/ is clearly longer, as can be confirmed by checking the /maka/-/paka/ comparison in Figure 5b. This difference in duration, we argue, is the result of two overlapping labial gestures. To see

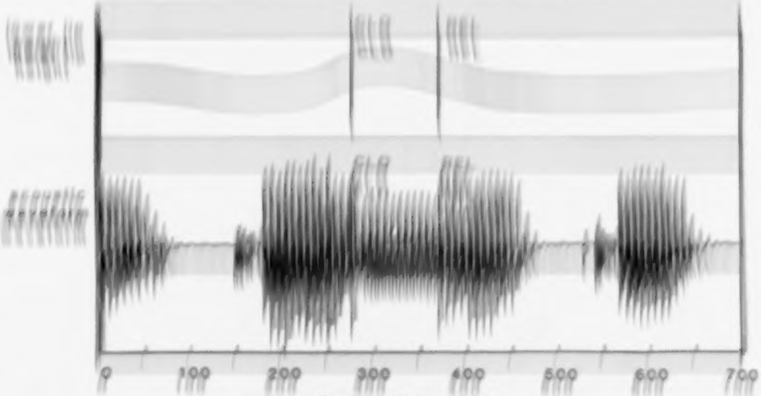


a. two superimposed tokens of English 'capper'

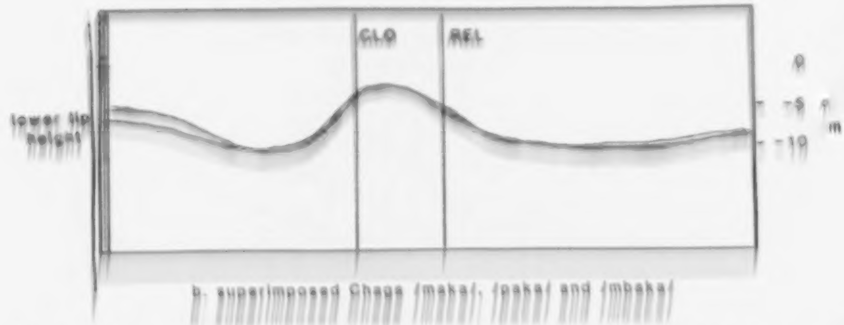


b. superimposed English 'capper,' 'cabber,' 'cammer,' 'camper,' and 'camber.'

Figure 3: Comparison of lower lip trajectories for English words.



a. two superimposed tokens of Chaga /maka/



b. superimposed Chaga /maka/, /paka/ and /mbaka/

Figure 4. Comparison of lower lip trajectories for Chaga words (except /mpaka/).

that this might be the case, consider Figure 5c, in which the gestures for /maka/ and /paka/ are superimposed in an overlapping fashion, and Figure 5d, in which the gesture for /mpaka/ is superimposed on the overlapping gestures for /maka/ and /paka/. The close correspondence observable in the figure holds across all the repetitions for these utterances. That is, the gesture for syllabic /m/ plus /p/ corresponds closely to the individual gestures for /m/ and /p/ arranged sequentially with partial overlap. An alternative description could be suggested, namely that the bilabial closure gesture in /mp/ was simply "larger." Note, however, that both the amplitude and the slopes of the onset and offset are unchanged from the single consonant case. This argues for overlap, or else another mechanism that simply holds the peak of a gesture, rather than for a larger gesture, since, as reported by Kelso, Vatikiotis-Bateson, Saltzman, and Kay (1985), larger gestures (due to changes in stress and rate, at least) typically have both increased amplitude and steeper slopes. We provisionally prefer the analysis of overlap to that of a held peak, since it requires mechanisms that must in any case occur in the phonology, namely overlap among gestures involving different articulators.

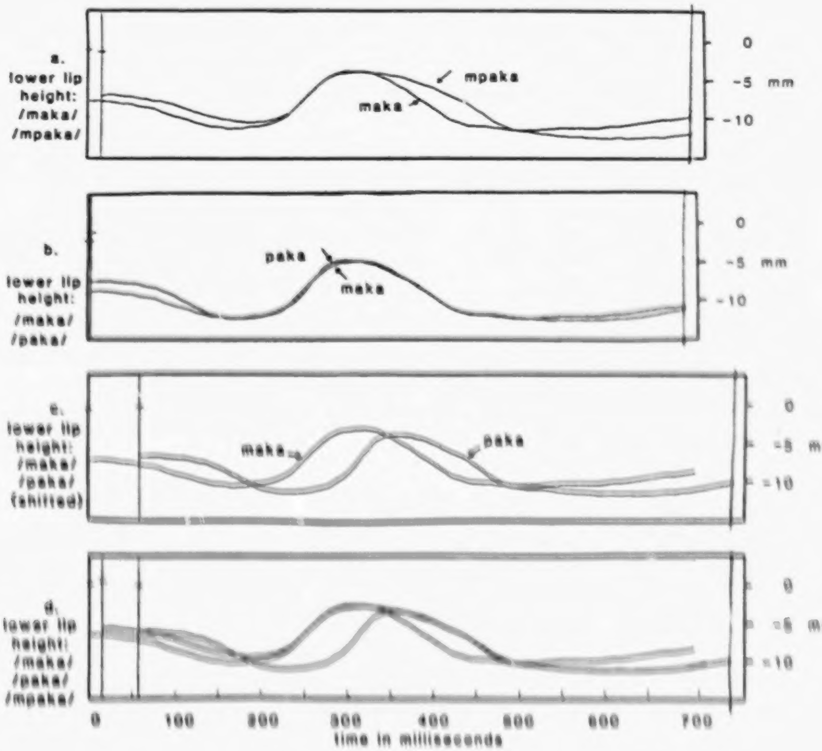


Figure 5. Comparison of lower lip trajectories for Chaga /mp/, /m/, and /p/.

Thus, the articulatory evidence suggests that syllabic /mp/ is a gestural constellation including two partially overlapping labial gestures. This distinguishes it from Chaga prenasalized stops and English nasal-stop clusters, both of which are constellations involving a single bilabial closure gesture. That is, as we hypothesized, the English nasal-stop clusters and Chaga prenasalized stops would both, using Anderson's (1976) framework, be represented as (2a). Representation (2b), however, would be more appropriate for the Chaga syllabic /mp/.

How then, given the similarity between their gestural structures, do we capture the distinction between prenasalized stops in Chaga and nasal-stop sequences in English? The simplest statement is as a distributional, or phonotactic, difference. That is, in Chaga such gestural structures can occur in word (and/or syllable) initial position, whereas in English the same gestural structures cannot occur in initial position. Thus, we can account for the difference between prenasalized stops and nasal-stop clusters as different distributional characteristics of the same physical structure. However, such a distributional difference can only serve to distinguish prenasalized stops and nasal-stop clusters in two different languages. We still need to address the issue of how prenasalized stops differ from nasal-stop clusters in a language where they contrast. In such a case, we expect an articulatory difference between the two.

Feinstein (1979) describes one such language, Sinhalese. We know that there is in fact a physical difference here between the nasal-stop clusters and prenasalized stops: Feinstein reports that the nasal is longer in /nd/ clusters than in prenasalized /^{n}d/. This difference might reside either in the relative timing of the oral and velic gestures, or in the oral closure gesture itself, which could be longer, or doubled, for the clusters. We would in fact expect the latter to be true, because the nasal-stop clusters are part of a morphological class of inanimate nouns in which gemination is an active process. That is, members of this class containing oral stops alternate between single and geminate stops in the plural and definite singular (e.g., [potu] and [pottə] 'core', Feinstein, 1979), while members containing nasal-stop sequences alternate between prenasalized and nasal-stop clusters (e.g., [ka^{n}du] and [kandə] 'hill', Feinstein, 1979). Such a classification would be simply explained if the identifying characteristic of the class were the lengthening, or doubling, of the oral gesture. For the oral stops, this would result in a geminate, while for the prenasalized stops, it would result in a lengthened nasal, i.e., a nasal-stop cluster.

Such a characterization, combined with a gestural reformulation of their syllable template, also directly captures Feinstein's (1979) and Cairns and Feinstein's (1982) analysis of the difference between the prenasalized and nasal-stop sequences in Sinhalese as a difference in syllable structure, with the prenasalized stops being tautosyllabic (syllable initial) and the nasal-stop sequences being heterosyllabic. In the gestural reformulation, the terminal nodes of the syllable template would be oral gestures, rather than segments, and syllable onsets would be restricted to single oral gestures. Since the velic gesture under this formulation is not relevant to the syllable structure, either prenasalized or single oral stops (assuming both have a single oral gesture) could occur in the syllable onset. That is, lengthened/doubled oral gestures would be heterosyllabic, while single oral gestures would be syllable initial, regardless of their cooccurrence with velic gestures. Such a reformulation correctly captures the syllabification difference in the nasal-stop sequences, and eliminates the need for a separate language-specific statement about the priority of prenasalization.

As we have shown, then, the gestural structures for prenasalized stops and nasal stop clusters lead to both a simple statement of their physical properties, and a satisfactory description of their phonological properties.

Anderson's (1976) analysis of prenasalized stops predicts their temporal properties fairly well, but for nasal-stop sequences in English, some rule would be required to collapse a structure like (2b) into a structure like (2a). Moreover, the structure in (2a) then must be mapped onto an articulatory representation, much like that which we take as our basic phonological representation. In contrast, in an articulatory phonology, the representation in terms of a spatio-temporal organization of gestures directly captures the articulatory structure as well as providing simple statements of phonological generalizations.

3. Preliminary Formalisms

The analyses of section 2 suggest that spatio-temporal descriptions of articulatory movements can, in fact, provide the basis for stating phonological regularities. In order to state such generalizations explicitly, the gestural structures must be formalized in some way, and it is to such formalizations that we turn in this section. We should note that these suggestions for formalization are preliminary and incomplete, and their implications for the description of a wide range of phonological data have not yet been investigated. Our preliminary attempts here are intended to simply show that the kind of structures that we have been arguing for can be rigorously formalized and that their adequacy in accounting for complex phonological data can, therefore, be tested.

We begin by describing one promising approach to a dynamical specification of gestures and their coordination in section 3.1. This dynamical specification is meant to be detailed enough to account for phenomena such as coarticulation and language-particular timing patterns. In section 3.2, we show that a simplified, more qualitative notation can be used to index these dynamically-defined structures. These indices are a simplified way of representing gestural structures, appropriate to such linguistic functions as lexical contrast and description of phonological generalizations.

3.1 Specification of Gestures and Inter-gestural Relations

We have been using the notion of an articulatory gesture as a characteristic pattern of movement of an articulator (or of an articulatory subsystem) through space, over time. How can we precisely define such spatio-temporal patterns? We might attempt to specify the values of articulator position at successive points in time. Such an approach, however, in which time is explicitly one dimension of the description and spatial position another, has trouble with the complex variations in articulatory trajectories introduced by changes in speaking rate and prosodic factors. It would be preferable to view the change in position over time as the output of a more abstract system, such as a dynamical system, capable of generating a variety of related trajectories.

Dynamical systems (e.g., Abraham & Shaw, 1982; Rosen, 1970) have been applied to problems of motor coordination in biological systems in general (e.g., Cooke, 1980; Fel'dman, 1966; Kelso & Holt, 1980; Kelso, Holt, Rubin, & Kugler, 1981; Kelso & Tuller, 1984a; Kugler et al., 1980; Polit & Bizzi, 1978) and to the organization of speech articulators in particular (Fowler, 1977; Fowler et al., 1980; Kelso & Tuller, 1984b; Kelso, Tuller, & Harris, 1983; Lindblom, 1967; Ostry & Munhall, 1985). Space and time are not specified in a point-by-point fashion in a dynamical system, but the system is capable of specifying characteristic patterns of movement that are organized in space and time. There are two properties of such systems that make them useful to the description of linguistic gestures. First, for a given fixed specification of system parameters, the system can output an infinite number of different (but related) trajectories, as a function of the initial conditions of the articulators, and as a function of other dynamical systems (for other gestures) that might be simultaneously active. At least some trajectory context-dependence (i.e., coarticulation) can be handled in this way, by characterizing a gesture in terms of the invariant input parameters to such a system. Second, although the articulators are moving throughout such a gesture, the equation itself does not vary over time, but rather characterizes the whole pattern of movement. Thus, the traditional notion of a discrete element, imposed on speech from the outside, is replaced by the notion of coherent gestural movements that can be described by a single system of equations (cf. Fowler et al., 1980).

To exemplify such equations, consider a physical example of a dynamical system, a mass attached to a spring. If the mass is pulled, stretching the spring beyond its rest length (equilibrium position), and then released, the system will begin to oscillate. Assuming that the system is without friction, the resulting movement trajectory of the mass can be described by the solution to equation (4).

(4) $m\ddot{x} + k(x - x_0) = 0$

where m = mass of the object
k = stiffness of the spring
x_0 = rest length of the spring (equilibrium position)
x = instantaneous position of the object
$\ddot{x}$ = instantaneous acceleration of the object

Thus, a time-varying trajectory is generated by an equation that does not itself change over time. Different trajectories can be obtained from this same system by different choices of values for the dynamical parameters m, k, and x_0, and by different initial conditions for x and $\ddot{x}$. Changing the stiffness k of the spring will affect the frequency of oscillation of the mass, while changing the rest length (equilibrium position) of the spring x_0 and the initial position x will affect the amplitude of oscillation.

Recently, it has been shown that such dynamical systems can account for systematic trajectory differences associated with linguistic variations in stress (Browman & Goldstein, 1985; Kelso et al., 1985; Ostry & Munhall, 1985). In these papers, mass-spring models such as (4) provided an abstract description of the articulatory movements associated with lip closure. Thus, the x in equation (4) was taken to represent the vertical position of the lower lip, instead of the length of a spring. The lower lip in the stressed syllables took more time to move between the displacement peaks and valleys, which was modelled by decreasing the stiffness (k) in equation (4). The lower lip in stressed syllables also moved a greater distance, which can be modelled by increasing the difference between the rest length of the lower lip (x_0) and the initial position (although this aspect of the modelling was less thoroughly explored in the above papers).

A gesture such as that for bilabial closure cannot be fully described by the movement of a single articulator; the coordination of a number of articulators is required, i.e., the jaw, the lower lip, and the upper lip. The *task dynamics* of Saltzman and Kelso (1983) offers a promising approach to modelling this coordination. Task dynamics provides an organization of articulatory movement that is defined in terms of a particular task to be performed--in our example, lip closure. It relates this task closure to the movement of the various articulators involved in its performance, in particular the jaw, the lower lip, and the upper lip. The positions of these articulators are anatomically linked--as the jaw moves, for example, the lower lip can move along with it. Because of this fact, lip closure can be achieved in a number of different ways, from moving only the jaw, to moving just the lower lip with relatively little jaw movement. It is this flexibility that allows a phonetic task such as bilabial closure to be achieved even when the movement of one of the component articulators is mechanically restrained during speech (Abbs, Gracco, & Cole, 1984; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984). Such compensatory behavior has been successfully modeled by the task dynamics approach (Saltzman forthcoming). In addition, this flexibility can account for aspects of coarticulation such as the fact that a bilabial closure gesture is produced with a higher jaw position in [bi] than in [bæ] (Sussman et al., 1973). The default contributions of the component articulators to a given task, in the absence of perturbation or coarticulation, can be specified in terms of characteristic weightings for these articulators. These weightings may vary for different gestures: for example, the upper lip is weighted quite differently in bilabial closure and labiodental fricative gestures.

A gesture, then, is defined by specifying (i) a dynamic equation (or a set of them), (ii) a motion variable or variables, i.e., the variable(s) to substitute for x in equation (4) or other dynamic equation, (iii) values for the coefficients of the equation (the dynamic parameters), and (iv) weightings for individual articulators. An initial application of this definition of gestures was presented in Browman, Goldstein, Kelso, Rubin, and Saltzman (1984). They employed a single undamped second order system (such as (4)) defined for two motion variables, lip aperture (vertical distance between the upper and lower lips) and lip protrusion. Different values for the dynamic parameters (stiffness and equilibrium position) were employed on alternate motion cycles, so as to generate the articulatory trajectories appropriate for an alternating stress sequence /'mama'mama..../. The computed output trajectories of the upper lip, lower lip, and jaw were then used to control a vocal-tract simulation (Rubin, Baer, & Mermelstein, 1981) that synthesized speech. Thus, a very simple specification of a bilabial closure gesture in terms of dynamically defined variables for lip aperture and protrusion successfully captured the information necessary to produce convincing speech. At the same time, as we have seen, such gestural descriptions are useful as a basis for phonological description.

In order to generate the complete inventory of speech sounds, gestures must be combined into constellations. Again, as with the gestures themselves, the relations among gestures can be specified abstractly using spatio-temporal phase relations (Kelso & Tuller, 1985). A specification in terms of phase is neither solely spatial nor solely temporal, because the exact point in time associated with a particular phase angle will change as the frequency (stiffness) of the gesture changes, and the exact point in space will change as the amplitude (rest length) of the gesture changes. Rather, phasing specifies relations among characteristic spatio-temporal patterns. Empirical evidence in favor of a spatio-temporal approach has been presented by Tuller, Kelso, and Harris (1982) and Tuller and Kelso (1984). For example, Tuller et al. (1982) have shown that in sequences like ['papip] and [pa'pip], the time of onset of lip activity for the intervocalic consonant relative to jaw or tongue activity for the initial vowel is quite variable across the two different stress patterns and across changes in speaking rate. This indicates that the purely temporal approach cannot specify gestural relations in a sufficiently general way. However, Tuller et al. go on to show that the onset of lip activity for the intervocalic consonant can be quite precisely (and linearly) related to the period of the vocalic cycle, defined as the time between the onset of activity for the first vowel and the onset of activity for the second. This linear relationship remains invariant across changes in speaking rate and stress. [Similar constancies in relative timing of acoustic events, across changes in rate, have been reported by Weismer and Fennell (1985).] Kelso and Tuller (1985) have further analyzed their movement data in terms of phase and have shown that the consonant gesture begins at a fixed phase angle in the vowel cycle. As the vowel period changes with stress and rate, the absolute time corresponding to that phase angle will also change, in a systematic way.

We take, then, as a first hypothesis that gestures can be characterized in terms of a dynamical system and its associated motion variables and parameter values, and that intergestural relations can be specified in terms of their phasing. This framework can accommodate the cross-linguistic timing differences discussed in section 1.0 quite naturally, although the analysis in any particular case (e.g., phase differences vs. stiffness differences) remains to be determined by the relevant spatial and temporal data.

3.2 Contrastive Articulatory Structures

So far, gestural organizations have been described in terms of attributes of the motions of physical articulators, including the more abstract and general physical descriptions provided by dynamics, and specifically task dynamics. That is, we have shown in the last section how it is possible to capture spatio-temporal articulatory structure, using a dynamical framework. In this section, we continue to develop a formalism for articulatory phonological representation by laying out some sample lexical descriptions.

One function of a lexical description is to provide information about the physical structure of an item so that linguistically significant similarities and differences among lexical entries can be observed in as simple and direct a way as possible. Since we are dealing solely with articulatory structure, the domain in which linguistic facts such as distinctiveness can be stated consists of the set of articulatory gestures and their relations. The descriptions in this section differ from those in the previous section only in terms of the degree of detail in the description. It is as if, in the present section, we have decreased the resolution on our microscope so that the descriptions are coarser-grained and more qualitative. Thus, we are referring to the same set of dynamically specified gestures, but this time using symbols which serve as indices to entire dynamical systems These symbolic descriptions highlight those aspects of the gestural structures that are relevant for contrast among lexical items. In our discussion of lexical representations, then, there are three important considerations to keep in mind. First, the minimal units in the lexical representation are dynamically-defined articulatory gestures. Second, these gestures are spatio-temporal in nature. Third, the gestures are organized asynchronously, with varying degrees of overlap among the gestures.

Table 1 shows the symbolic notation we are suggesting to index the gestures relevant to the English and Chaga words discussed in section 2.2. The symbols are shorthand for specific sets of dynamical equations and their associated motion variables and parameter values that can generate the kinds of articulatory trajectories seen in Figure 6. These gestural trajectories are based on the articulatory data for <camper>, an example of which was shown in Figure 2. The bottom panel is the acoustic signal. The middle articulatory panel is the actual recorded trace of the vertical movement of the lower lip. This closing and opening gesture of the lips is indexed by the β in Table 1. The other two panels of articulators are estimates only, and show the amount of opening associated with the gesture, rather than actual vertical height. The bottom articulatory panel displays a representative glottal gesture associated with voicelessness: the peak indicates the maximum opening of the vocal folds. [The shape of the glottal gesture was estimated from Sawashima & Hirose (1983), while the timing was based on the acoustic signal, combined with information from various studies on glottal timing (Löfqvist, 1980; Löfqvist & Yoshioka, 1985).] Here, a γ in Table 1 represents the glottal opening and closing gesture. Note that the presence of the glottal gesture means an open glottis, i.e., voicelessness.

The top panel estimates the opening of the velo-pharyngeal port, based on the accelerometer record and published data on velum movement (e.g., Kent, Carney, & Severeid, 1974; Vaissiere, 1981). These data indicate that in an utterance like <camper>, velum lowering (i.e., velic opening) begins at the release of the initial consonant, and velum raising (i.e., velic closing) begins at some time before the achievement of articulatory closure for the /mp/. The velum movement is represented by two separate gestures in Table 1, an opening gesture indicated by a +μ (nasal), and a closing gesture indicated by a -μ (non-nasal, i.e., oral). The decision to treat velic opening and closing as two separate gestures, as compared with the glottal and oral gestures that incorporate both opening and closing, is based on the fact that each velic gesture may act as a word-level phenomenon, so that the velum can possibly be held in either a closed or an open position indefinitely. Kent et al. (1974) provide an example of the latter phenomenon: a long sentence with many nasal consonants in which the velum remains lowered throughout. This can also be seen in nasal harmony as well as in non-distinctive nasalization. In addition, the velum opening and closing gestures may require different amounts of time; for example, Benguerel, Hirose, Sawashima, and Ushijima (1977) show that for a French talker, opening gestures are slower than closing gestures. It is possible that a comparable internal gestural structure will also be needed for oral and glottal gestures.

Figure 6 also illustrates the timing relations among gestures. Note that the overlap among the gestural trajectories reflects the inherent spatio-temporal properties of the gestures as well as their asynchronous organization. That is, the gestures that form the /mp/ constellation all require a certain amount of time to unfold, but they are not necessarily synchronized in the sense of their onsets (or peaks, or offsets) being aligned. The velic opening, for example, begins at least 100 ms before the labial gesture begins, and velic closure begins sometime in the middle of the labial gesture. This asynchrony results in the vowel being nasalized, and the labial gesture being partly nasal and partly oral. The glottal gesture is

Table 1

Gestural Symbols

symbol	gesture
β	bilabial closing and opening
γ	glottal opening and closing (returns to voicing position)
$^{+}\mu$	velic opening
$^{-}\mu$	velic closing
V	vowel

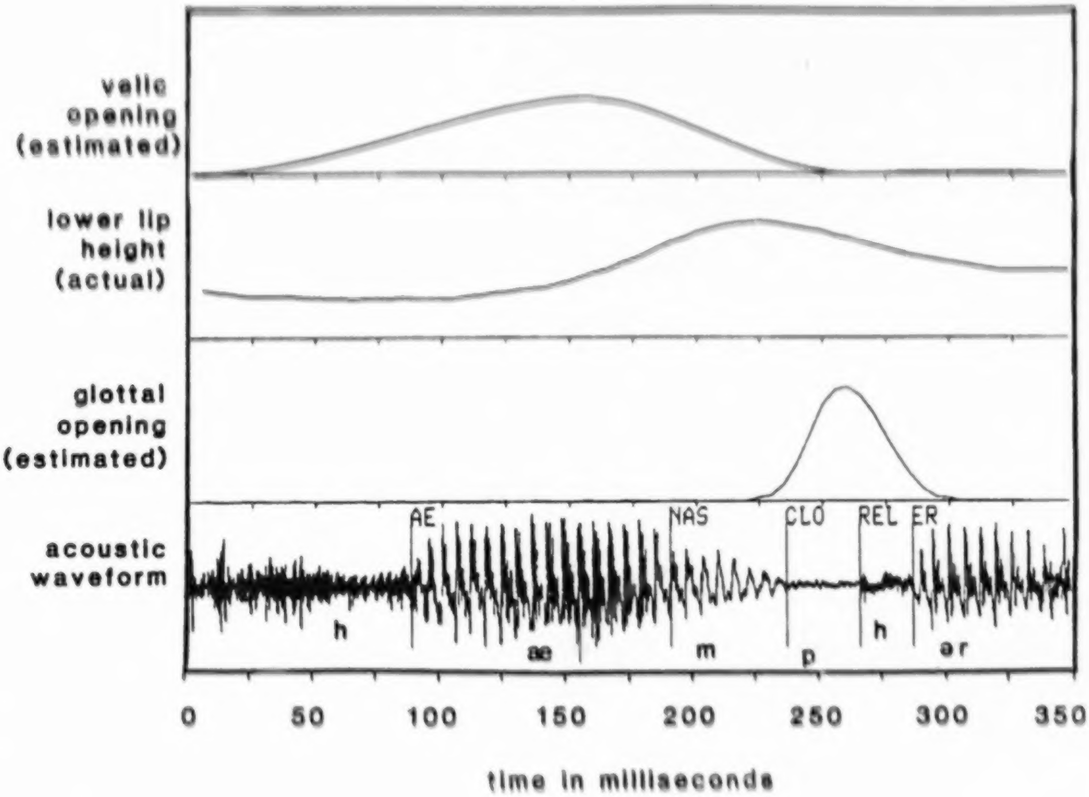


Figure 6. Acoustic waveform and bilabial, velic, and glottal gestures for <camper>.

delayed with respect to the labial gesture, so that the glottis most likely reaches its peak opening after the peak closure of the labial gesture, but before its offset. Thus, a single glottal gesture, requiring a certain amount of time to unfold and asynchronously aligned with the labial closure gesture, generates both voicelessness and aspiration (as originally proposed by Lisker and Abramson 1964). In addition to the particular gestures, then, our symbolic representation must capture aspects of the phase relations among the gestures, because as we shall see, contrasting items may include the same set of gestures, but in different relations.

For the sake of our symbolic representations, we project gestural constellations into a two-dimensional representation which captures qualitative aspects of these relations that are important for contrast (or for certain kinds of phonological generalizations). Examples of such constellation projections for the English words investigated in section 2.2 are shown in Figure 7, with gestures for the initial consonant omitted. Chaga words beginning with /m/, /p/, and /mb/ have representations like those of the comparable English words, except that the initial V is not present for these words. The representation of Chaga /mpaka/ (with syllabic /m/) is also shown in the figure. Note that vocalic gestures are indexed simply as V, because their gestural aspects have not been investigated here.

The vertical dimension of these representations is organized by articulatory subsystem. Gestures of the oral subsystem are found on the top two lines, gestures of the laryngeal subsystem are found on the third line, and velic gestures are at the bottom. The particular ordering (from top to bottom) is meant to relate the gestural structures to the more global (syllable and foot) rhythmic organization of speech. (This rhythmic organization, corresponding to, e.g., metrical trees or grids, or CV skeleta, is itself not yet incorporated in these structures). The closer to the top a gesture is, the more relevant it is presumed to be in carrying the overall rhythm. Thus, vowel gestures are found on the top line, as they seem to be most important in carrying the speech rhythm, with other gestures being coproduced with them (cf. Fowler, 1983). Velic gestures, by contrast, are placed at the very bottom, because they contribute very little, by themselves, to the rhythmic structure.

The horizontal dimension of the representations in Figure 7 consists of a grid that can be used to give a qualitative indication of the relative phase relations of the gestures. The lines of the grid represent roughly 90 degree (quarter cycle) phase intervals. Two gestures that are lined up on the same grid line are assumed to be relatively synchronous. For example, their onsets, or their maximum displacements, might coincide in time. The grid lines are not, however, meant to indicate any special structural relation between lined-up gestures. For example, it is not necessarily the case that one of the gestures governs the other, or that there is a particularly cohesive (or tightly invariant) relationship between the gestures. Gestures on successive grid lines are assumed to have approximately a 90 degree phase relation (e.g., the displacement maximum of one gesture synchronized with the velocity maximum of another); those two lines apart are assumed to have a 180 degree relation; etc.

As an example, consider the constellation for English <camper>. The placement of the two V symbols indicates that they are phased 360 degrees apart (they are separated by four grid intervals), and thus, they represent one complete vowel cycle. For expository purposes, we can think of this vowel

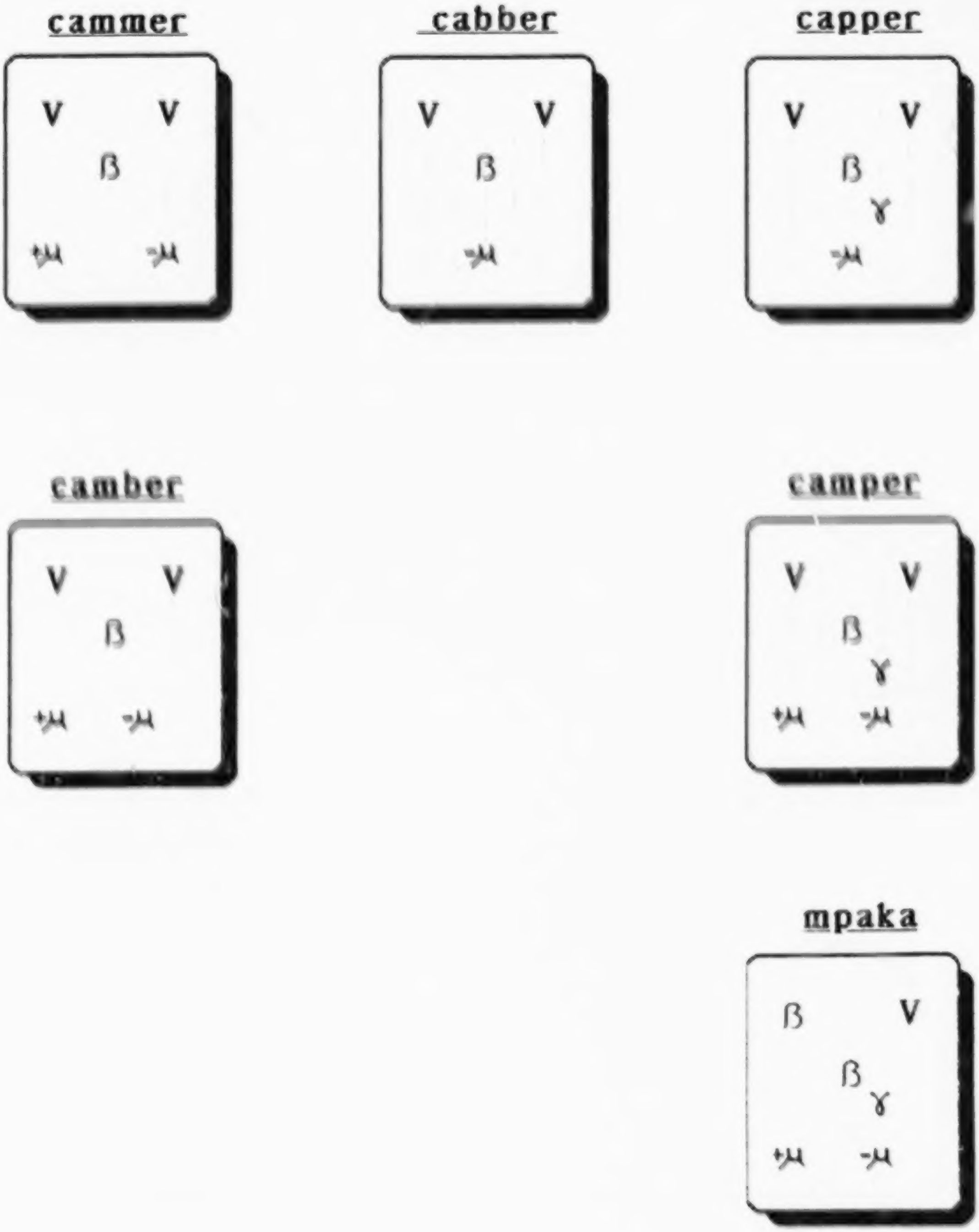

cammer
V V
β
+μ -μ
cabber
V V
β
-μ
capper
V V
β
γ
-μ
camber
V V
β
+μ -μ
camper
V V
β
γ
+μ -μ
mpaka
β V
β
γ
+μ -μ

Figure 7. Gestural constellations for English words, and for Chaga /mpaka/ (see text for interpretation).

cycle in terms of the action of the jaw--high for the consonant, low for the vowel. The first grid line can be thought of as corresponding to the onset of the first vowel gesture--the beginning of the descent of the jaw from its maximum height towards its low position for the vowel. The third grid line can then be used for gestures that are 180 degrees out of phase with respect to the vowel gesture, i.e., a gesture whose onset occurs 180 degrees into the vowel cycle, when the minimum jaw height is reached for the vowel. This is (roughly) the phase relation between the bilabial closure gesture (on the third grid line) and the vowel gestures reported by Kelso and Tuller (1985). Note that this representation also directly captures the notion (Öhman, 1966; Fowler, 1983) that consonant gestures overlap a continuous vowel production cycle.

The glottal gesture in <camper> is positioned 90 degrees out of phase with [illegible]

Such representations not only provide qualitative information about the articulatory structure for individual items, but also serve to differentiate items from each other. Compare, for example, the representations for English <camper> and Chaga /mpaka/. Here, based on our articulatory measurements, the distinction between the English /mp/ and the Chaga syllabic nasal+stop is represented by a second labial gesture for the Chaga. This gesture is positioned on the top line (where normally only vowel gestures occur), in order to capture the fact that this bilabial closure gesture assumes the syllabic function within rhythmic structure that is more usually filled by vowel gestures.

Another pair of lexical items, <cammer> and <camber>, demonstrates the distinctive use of phase structure in the representations. The only difference between the representations for <cammer> and <camber> lies in the phasing of the velic closing gesture. In <camber>, the velum closes during the labial gesture, so that there is a brief period of non-nasalized closure. In <cammer>, on the other hand, the velum closes sometime after the labial gesture is released. This is captured by different grid positions of the velic closure gesture. (The phasing of the gestures is inferred from evidence provided by the nasal accelerometer combined with the acoustic signal.)

It is important to note that the physical descriptions provided by these lexical entries are not complete descriptions of the articulatory actions. Other physical or physiological events regularly occur in these utterances but are not part of these descriptions. For example, larynx lowering and overall expansion of the oral cavity (Westbury, 1983) typically accompany the bilabial closure for voiced stops. However, at this early stage of our investigation, we wish to focus on characterizing those aspects of the physical structure that are most relevant to capturing linguistic generalizations and specifying contrast among lexical items. In addition, these qualitative representations may omit certain detailed differences that are present in the quantitative specification of the gestures' dynamic parameter values and phasing. For example, the phase angle for /b/ relative to the vowel is somewhat greater than it is for /p/ (approximately 205 degrees vs. 180 degrees in Kelso & Tuller, 1985). This difference corresponds to the durational differences (discussed in section 1) between vowels before voiced and voiceless stops. Such differences in gestural parameters and phasing are directly represented in the more quantitative description.

The gestural constellations in Figure 7 represent contrast in a physically realistic way; however, the representations are clearly understructured when compared to recent forms of phonological representation. We expect, however, that additional structuring will emerge as we learn more about patterns of intergestural phasing, [illegible] in terms of the organization [illegible]

The phase relations among gestures are reminiscent of association lines among [illegible] phasing and autosegmental associations share the advantage of allowing gestural overlap (in gestural terms) or multiple associations among autosegments (in autosegmental terms). From this perspective, an articulatory phonology and autosegmental phonology can be seen as converging on the same type of lexical representation. There is nothing in a gestural framework that contradicts autosegmental representations. Rather, autosegmental phonology and the present framework differ first in their starting points (phonological patterns vs. articulatory measurements), and second in the aspects of the representation that are more highly structured. In particular, the gestures have an explicit internal structure: they are dynamical systems that serve to structure the movements of articulatory subsystems.

Thus, the gestural framework can provide a basis for making some principled predictions about the likelihood of a particular phonological feature (or set of features) being split off as a separate tier in autosegmental phonology. We would expect, in general, that the more articulatorily independent of other features a particular feature is, the more likely it is to become a separate autosegmental tier. In particular, we would predict that phonological features associated with gestures of the three articulatory subsystems would be the most likely to be segregated onto separate tiers, a view that is compatible with the prevalence of autosegmental analyses for nasalization and tone. Within the oral subsystem, the gestures involving the lips are relatively independent of the tongue gestures (although both sets share the jaw as an articulator). Thus, features involving lip gestures would be the next most likely to be segregated onto a tier. More generally, we expect that successive gestures specified with common motion variables but with different dynamic parameter values would be more likely to be grouped together on a separate tier than gestures with similar parameter values but specified for different motion variables. Thus, we expect that place of articulation features (which correspond to motion variables) should readily split off onto separate tiers but that manner or constriction degree features like [continuant] or [high] (which correspond to particular values of the dynamic parameters) should do so more rarely. Formally speaking, there is no distinction in traditional feature systems between the features for place and manner, whereas in the gestural analysis, they correspond to distinct aspects of the representation. In general then, the explicit model of articulatory organization provided by the gestural model can lead to specific hypotheses about the hierarchy of expected tier independence.

4. Concluding Remarks

We have outlined an approach to phonological representation based on [illegible] that can be accounted for using gestural structures and tiers.

Even within other phonological approaches, a gestural description of speech could be used as the basic data for which the phonology attempts to account. That is, gestural structures could replace phonetic transcriptions as the "output" of the phonology. There are several reasons for doing so. First, it is easier to verify empirically the gestural structure of an utterance: the relation between gestural structures and physical observables is simple and constrained, compared to the relation between a segmental transcription and speech. Second, the gestural structure incorporates temporal information that is usually absent from segmental transcriptions. This is not only important for its own sake (in accounting for cross-language differences, for example), but the increased resolution of the representation may sharpen the process of comparing competing phonological analyses. Finally, as we have argued above, certain aspects of phonological representations, such as tier decomposition, can be rationalized or explained with respect to such gestural structures.

References

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming. Journal of Motor Behavior, 16, 195-231.

Abercrombie, D. (1967). Elements of general phonetics. Edinburgh: Edinburgh University Press.

Abraham, R. H., & Shaw, C. D. (1982). Dynamics-The geometry of behavior. Santa Cruz, CA: Aerial Press.

Anderson, J., & Jones, C. (1974). Three theses concerning phonological representations. Journal of Linguistics, 10, 1-26.

Anderson, S. R. (1974). The organization of phonology. New York: Academic Press.

Anderson, S. R. (1976). Nasal consonants and the internal structure of segments. Language, 52, 326-344.

Anderson, S. R. (1978). Syllables, segments, and the Northwest Caucasian languages. In A. Bell & J. B. Hooper (Eds.), Syllables and segments. Amsterdam: North-Holland.

Aronoff, M., & Oehrle, R. T. (Eds.). (1984). Language sound structure. Cambridge, MA: MIT Press.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. Phonetica, 38, 9-20.

Benguerel, A. P., Hirose, H., Sawashima, M., & Ushijima, T. (1977). Velar coarticulation in French: A fiberscopic study. Journal of Phonetics, 5, 149-158.

Bernstein, N. (1967). The coordination and regulation of movements. London: Pergamon Press.

Browman, C. P., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), Phonetic linguistics (pp. 35-53). New York: Academic.

Browman, C. P., Goldstein, L., Kelso, J. A. S., Rubin, P., & Saltzman, E. (1984). Articulatory synthesis from underlying dynamics. Journal of the Acoustical Society of America, 75, [illegible]

[illegible] e theory of [illegible]

[illegible] Fundamental problems in phonetics. [illegible]nington, IN: [illegible]

[illegible] The sound pattern of English. New York: [illegible]

[illegible] Vowel harmony in nonlinear generative phonology: An [illegible] Indiana University Linguistics Club.

Clements, G. N., & Keyser, S. J. (1983). CV phonology: A generative theory of the syllable. Cambridge, MA: MIT Press.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 199-212). Amsterdam: North-Holland.

Ewen, C. J. (1982). The internal structure of complex segments. In H. van der Hulst & N. Smith (Eds.), The structure of phonological representations (Part II, pp. 27-67). Cinnaminson, NJ: Foris Publications U.S.A.

Feinstein, M. H. (1979). Prenasalization and syllable structure. Linguistic Inquiry, 10, 245-278.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic analysis of execution by man of the simplest motor tasks. Biophysics, 11, 766-775.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. Language and Speech, 24, 125-146.

Fourakis, M. S. (1980). A phonetic study of sonorant-fricative clusters in two dialects of English. Research in Phonetics (Department of Linguistics, Indiana University), 1, 167-200.

Fowler, C. A. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. Journal of Phonetics, 8, 113-133.

Fowler, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*, 386-412.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 373-420). New York: Academic.

Fromkin, V. A. (Ed.). (1985). *Phonetic linguistics*. New York: Academic Press.

Fujimura, O. (1981). Temporal organization of articulatory movements as a multidimensional phrasal structure. *Phonetica*, *38*, 66-83.

Fujimura, O., Kiritani, S., & Ishida, H. (1973). Computer controlled radiography for observation of movements of articulatory and other human organs. *Computers in Biology and Medicine*, *3*, 371-384.

Fukui, N., & Hirose, H. (1983). Laryngeal adjustments in Danish voiceless obstruent production. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, Tokyo), *17*, 61-71.

Goldsmith, J. A. (1976). *Autosegmental phonology*. Bloomington, IN: Indiana University Linguistics Club.

Haggard, M. (1973). Abbreviation of consonants in English pre- and post-vocalic clusters. *Journal of Phonetics*, *1*, 9-24.

Halle, M., & Vergnaud, J. R. (1980). Three dimensional phonology. *Journal of Linguistic Research*, *1*, 83-105.

Hayes, B. (1981). *A metrical theory of stress rules*. Bloomington, IN: Indiana University Linguistics Club.

Herbert, R. K. (1975). Reanalyzing prenasalized consonants. *Studies in African Linguistics*, *6*, 105-123.

Hockett, C. F. (1955). *A manual of phonology*. Baltimore, MD: Waverly Press.

Hooper, J. B. (1972). The syllable in phonological theory. *Language*, *28*, 525-540.

Hooper, J. B. (1976). *An introduction to natural generative phonology*. New York: Academic Press.

Jespersen, O. (1914). *Lehrbuch der Phonetik*. Leipzig: B. G. Teubner.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Bloomington, IN: Indiana University Linguistics Club.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*, 286-319.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin (Ed.), *Phonetic linguistics*. New York: Academic Press.

Kelso, J. A. S., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. *Journal of Neurophysiology*, *43*, 1183-1196.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. *Journal of Motor Behavior*, *13*, 226-261.

Kelso, J. A. S., & Tuller, B. (1984a). A dynamical basis for action systems. In M. Gazzaniga (Ed.), *Handbook of cognitive neuroscience* (pp. 321-356). New York: Plenum.

Kelso, J. A. S., & Tuller, B. (1984b). Converging evidence in support of common dynamical principles for speech and movement coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, *246*, R928-R935.

Kelso, J. A. S., & Tuller, B. (1985). Intrinsic time in speech production: Theory, methodology, and preliminary observations. *Haskins Laboratories Status Report on Speech Research*, *SR-81*, 23-39.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech (pp. 137-173). New York: Springer-Verlag.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Kent, R. D., Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. Journal of Speech and Hearing Research, 17, 470-488.

Kent, R. D., & Moll, K. L. (1969). Vocal-tract characteristics of the stop cognates. Journal of the Acoustical Society of America, 46, 1549-1555.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Kuipers, A. H. (1976). Typologically salient features of some Northwest Caucasian languages. Studia Caucasia, 3, 101-127.

Ladefoged, P. (1971). Preliminaries to linguistic phonetics. Chicago: Chicago University Press.

Ladefoged, P. (1980). What are linguistic sounds made of? Language, 56, 485-502.

Lass, R. (1984). Phonology: An introduction to basic concepts. Cambridge: Cambridge University Press.

Lehiste, I. (1970). Suprasegmentals. Cambridge, MA: MIT Press.

Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-436.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), Language sound structure (pp. 157-233). Cambridge, MA: MIT Press.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. Linguistic Inquiry, 8, 249-336.

Lindau, M. (1984). Phonetic differences in glottalic consonants. Journal of Phonetics, 12, 147-155.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. Quarterly Progress and Status Report (Speech Transmission Laboratory, University of Stockholm), QPSR-4, 1-29.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. Papers from the Institute of Linguistics (University of Stockholm), 21, 1-58.

Lisker, L. (1974). On time and timing in speech. In T. Sebeok (Ed.), Current trends in linguistics (pp. 2387-2418). The Hague: Mouton.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384-422.

Löfqvist, A. (1980). Interarticulator programming in stop production. Journal of Phonetics, 8, 475-490.

Löfqvist, A., & Yoshioka, H. (1980b). Laryngeal activity in Swedish obstruent clusters. Journal of the Acoustical Society of America, 68, 792-801.

Löfqvist, A., & Yoshioka, H. (1980). Laryngeal activity in Icelandic obstruent production. *Haskins Laboratories Status Report on Speech Research*, *SR-63/64*, 272-292.

Löfqvist, A., & Yoshioka, H. (1985). Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication*, *3*, 279-289.

Lovins, J. B. (1978). 'Nasal reduction' in English syllable codas. *Chicago Linguistic Society*, *14*, 241-253.

MacNeilage, P. F. (Ed.). (1983). *The production of speech*. New York: Springer-Verlag.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 415-420). Amsterdam: North-Holland.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, *12*, 373-418.

McCarthy, J. J. (1984). Prosodic organization in morphology. In M. Aronoff, R. T. Oehrle, F. Kelley, & B. W. Stephens (Eds.), *Language sound structure* (pp. 299-317). Cambridge, MA: The MIT Press.

Mitleb, F. M. (1984). Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics*, *12*, 23-27.

Nurse, D. (1979). *Classification of the Chaga dialects*. Hamburg: Helmut Buske Verlag.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (Eds.), *Historical linguistics II: Theory and description in phonology* (pp. 352-389). Amsterdam: North Holland.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*, 151-168.

Ostry, D. J., & Munhall, K. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America*, *77*, 640-648.

Petursson, M. (1977). Timing of glottal events in the production of aspiration after [s]. *Journal of Phonetics*, *5*, 205-212.

Pike, K. L. (1943). *Phonetics*. Ann Arbor, MI: University of Michigan Press.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. *Science*, *201*, 1235-1237.

Port, R. F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, *69*, 262-274.

Port, R. F., & O'Dell, M. L. (1984). Neutralization of syllable final voicing in German. *Research in Phonetics* (Department of Linguistics, Indiana University), *4*, 93-133.

Prince, A. S. (1984). Phonology with tiers. In M. Aronoff, R. T. Oehrle, F. Kelley, & B. W. Stephens (Eds.), *Language sound structure* (pp. 234-244). Cambridge, MA: The MIT Press.

Raphael, L. J., Dorman, M. F., Freeman, F., & Tobin, C. (1975). Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies. *Journal of Speech and Hearing Research*, *18*, 389-400.

Rosen, R. (1970). *Dynamical system theory in biology. Vol. I: Stability theory and its applications*. New York: WileyInterscience.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, *70*, 321-328.

Saltzman, E. (in press). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental Brain Research Supplementum*.

Saltzman, E., & Kelso, J. A. S. (1983). Skilled actions: A task dynamic approach. *Haskins Laboratories Status Report on Speech Research, SR-76*, 3-50.

Sawashima, M., & Hirose, H. (1983). Laryngeal gestures in speech production. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 11-38). New York: Springer-Verlag.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, *11*, 563-605.

Stelmach, G. E., & Requin, J. (Eds.). *Tutorials in motor behavior*. Amsterdam: North-Holland.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, *16*, 397-420.

Tuller, B., & Kelso, J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariants. *Journal of the Acoustical Society of America*, *76*, 1030-1036.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 460-472.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing* (pp. 211-265). Hillsdale, NJ: LEA.

Vaissiere, J. (1981). Prediction of articulatory movement from phonetic input. *Journal of the Acoustical Society of America*, *70*, S14 (A).

Vatikiotis-Bateson, E. (1984). The temporal effects of homorganic medial nasal clusters. *Research in Phonetics* (Department of Linguistics, Indiana University), *4*, 197-233.

Walsh, T., & Parker, F. (1982). Consonant cluster abbreviation: An abstract analysis. *Journal of Phonetics*, *10*, 423-437.

Weismer, G., & Fennell, A. M. (1985). Constancy of (acoustic) relative timing measures in phrase-level utterances. *Journal of the Acoustical Society of America*, *78*, 49-57.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, *73*, 1322-1326.

Yoshioka, H., Löfqvist, A., & Hirose, H. (1981). Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America*, *70*, 1615-1623.

Footnotes

[1] The notion of a gesture has been used in a somewhat similar way as a basic phonological unit in recent versions of dependency phonology (Ewen, 1982; Lass, 1984). However, our use of the term gesture is restricted to patterns of articulatory movement, while in dependency phonology it can refer to units that cannot, in any obvious sense, be defined in that way, such as the "categorial" gestures for major classes.

[2] It might be possible to link the two statements by means of markedness conventions. Keating (1984), reanalyzing Trubetzkoy's markedness theory, has argued that voiceless unaspirated stops tend to appear in various languages in positions of neutralization.

REPRESENTATION OF VOICING CONTRASTS USING ARTICULATORY GESTURES*

Louis Goldstein† and Catherine P. Browman

The representation of cross-language voicing contrasts has been a recurrent problem, since the mapping between phonological categories and their physical phonetic realizations is not one-to-one. Recently, Keating (1984) has argued that the representation of such contrasts for stop consonants must involve purely abstract features ([+voice] and [-voice]), which map onto phonetic categories for stops based on voice onset time (voiced, voiceless unaspirated, voiceless aspirated) in different ways for different languages. However, an articulatory analysis of voicing contrasts based on the presence or absence of glottal opening-and-closing gestures, as suggested in Browman and Goldstein (1986), may well provide a more nearly one-to-one mapping between phonological and physical categories. Moreover, as we shall show, such an articulatory analysis correctly predicts patterns of F_0 behavior that are wrongly predicted on the basis of purely abstract voicing categories.

Keating (1984) argues that if phonological features are constrained to be the same features as those used for phonetic representation, then certain cross-linguistic generalizations, involving voicing assimilation and correlations of voicing with vowel duration and pitch, will be missed. She demonstrates that the phonetic classes of voiced, voiceless unaspirated, and voiceless aspirated do not provide adequate cross-language natural classes for phonological rules. For example, both French and English have a voicing contrast, but different phonetic categories are involved. Whereas French (and sometimes English, in utterance-medial position) contrasts fully voiced [b,d,g] with voiceless unaspirated [p,t,k], English can contrast voiceless unaspirated [p,t,k] with voiceless aspirated [p^h,t^h,k^h] (in absolute initial position). Thus, the phonetic categories that contrast are not the same in the two languages. Nevertheless, as with many other languages, the vowels in both French and English are longer before the phonologically [+voice] stops than before the phonologically [-voice] stops (cf. Mack, 1982). Similarly, cluster voicing assimilation is found in languages regardless of whether the stops contrast in voicing or aspiration.

We suggest that Keating's abandonment of physically-based phonological features may in fact be unnecessary, and may simply reflect the wrong choice of physical descriptors. That is, a description in terms of articulatory gestures and their relative timing is, in fact, capable of accounting for the patterns Keating discusses. The basic phonological units in such an approach

---

*Journal of Phonetics, in press.

†Also Department of Linguistics, Yale University

Acknowledgment. This work was supported in part by NIH grants HD-01994, NS-13870, NS-13617 to Haskins Laboratories.

are articulatory gestures: organized patterns of movement within the oral, laryngeal, and nasal articulatory systems. Thus, as formalized in Browman and Goldstein (1986), voiceless stops can, to a first approximation, be represented as constellations of two gestures (an oral constriction gesture tightly coordinated with a glottal opening-and-closing gesture), while voiced stops can be represented as single oral constriction gestures. As originally suggested by Lisker and Abramson (1964), differences between aspirated and unaspirated voiceless stops can be captured directly by the timing between the two gestures in the constellation (or phasing: cf. Kelso & Tuller, 1985; Browman & Goldstein, 1986).

Voicing contrasts. In a gestural approach, the voicing contrast in both French and English is described as the presence vs. the absence of a glottal opening-and-closing gesture. In utterance-medial position, both French and English [-voice] stops typically show glottal opening-and-closing gestures, regardless of whether they are unaspirated (French, English) or aspirated (English). The [+voice] stops, however, do not display glottal opening-and-closing gestures in either language. This correlation between contrastive voicing and the presence vs. absence of glottal gestures can be seen for French in the data of Benguerel, Hirose, Sawashima and Ushijima (1978). For English, Lisker, Abramson, Cooper, and Schvey (1969) found that in running speech 96 percent of stressed /ptk/, 84 percent of unstressed /ptk/, and only 6 percent of /bdg/ were produced with glottal opening-and-closing gestures. Although the timing and size of the glottal gesture in English and French differ, the categorization of stops as [-voice] or [+voice] in utterance-medial position correlates quite well in both languages with the presence vs. absence of a glottal opening-and-closing gesture.

In absolute initial position, the glottis is already open (for breathing), and the opening portion of the glottal opening-and-closing gesture is, therefore, not actually observed. Thus, the relevant difference between [+voice] and [-voice] stops in this position is in the relative timing of the adduction of the vocal folds. Both French /d/ (Benguerel et al., 1978) and English /b/ (Flege, 1982) show glottal adduction well before stop release in utterance-initial position. Note that this is true for English (for eight of the ten speakers) regardless of whether there is voicing during the closure. That is, both phonetically voiced and voiceless unaspirated /b/ can show the same pattern of glottal adduction. Thus, a physical characterization using articulatory gestures appears to capture the voicing contrast in English and French for utterance-initial as well as utterance-medial position.

Vowel length. The simplest, and strongest, claim in a gestural approach is that vowel length differences will be correlated with the absence (longer) and presence (shorter) of a glottal opening-and-closing gesture. For those languages on which both glottal and durational data are available, the strong claim appears to hold up. French and English, as discussed above, are clear examples. Dutch (Slis & Cohen, 1969), Swedish (Lindblom & Rapp, 1973), and Korean (Chen, 1970) all display the vowel length difference, and available data suggest that the contrast for these languages can be described as the presence vs. absence of glottal gestures: Dutch (Slis & Damste, 1967) and Swedish (Lindqvist, 1972) both contrast [-voice] stops that have glottal opening-and-closing gestures with [+voice] stops that do not. Korean presents a slightly more complicated case in utterance-initial position, but in the intervocalic environment relevant to the vowel length rule, the same pattern is found (Kagaya, 1974).

Voicing assimilation. In a gestural approach, assimilation of a cluster in favor of the voiceless member can be described as a rule specifying the overlap of a single glottal gesture with two oral gestures. Assimilation in favor of the voiced member would be described as glottal gesture deletion. Thus, the gestural approach can account for voicing assimilation rules quite naturally.

F_0 patterns. Keating (1984) also discusses the relation between voicing contrasts and F_0 patterns. In many languages, the F_0 pattern on vowels following [+voice] stops is low and rising, while that following [-voice] stops is high and falling. Keating presents evidence from Hombert, Ohala, and Ewan (1979) that the F_0 patterns on the vowels following stops in French and English are similar. The F_0 following /ptk/ in either language shows a high falling pattern, while following /bdg/ it shows a low rising pattern. The F_0 difference is thus seen by Keating as reflecting the underlying abstract status, rather than the phonetic realization, since in the data of Hombert et al. (1979), English /bdg/ and French /ptk/ fall together phonetically as voiceless unaspirated. However, we can once again associate the similar behavior of French and English with the fact that for both languages, the [-voice] stops have a glottal opening-and-closing gesture while the [+voice] ones do not.

The relation between voicing and F_0 in Danish provides the most interesting comparison of the abstract and gestural analyses. In Danish, there is a contrast in initial position between aspirated and unaspirated stops. Unlike other contrasts described so far, however, both stops show glottal opening gestures (Frøkjaer-Jensen, Ludvigsen, & Rischel, 1973). (The unaspirated stops have smaller glottal gestures and are timed differently.) Keating's abstract analysis predicts that Danish should behave like English and French in showing a high falling F_0 pattern following [-voice] stops and a low rising pattern following [+voice] stops, since all three languages contrast [±voice] stops. A gestural analysis predicts that, on the contrary, the Danish stops, both of which have glottal gestures, should both show high falling F_0 patterns. The gestural analysis, therefore, predicts that Danish will be unlike French and English, which contrast presence vs. absence of glottal gestures. Petersen's (1983) study of F_0 following initial consonants in Danish supports the prediction based on the gestural analysis. The F_0 patterns following aspirated and unaspirated stops are the same--high and falling (with a pitch difference averaging only 2 Hz). Moreover, the Danish consonants examined that do not have glottal opening-and-closing gestures (/v/ and /m/) show a low rising F_0 pattern. While these results must be treated with some caution, as other studies of Danish have revealed larger F_0 differences between aspirated and unaspirated stops (Jeel, 1975), nevertheless Petersen's study provides clear evidence for a correlation between glottal gestures and F_0 patterns, rather than between abstract voicing categories and F_0 patterns.

Thus, analysis of cross-linguistic voicing contrasts in terms of glottal opening-and-closing gestures accounts for the similarities between languages as well as or, in the case of F_0 patterns, better than the purely abstract analysis posited by Keating. In addition, the articulatory analysis captures the facts of articulation directly, rather than requiring an additional set of mapping functions.

References

Benguerel, A. P., Hirose, H., Sawashima, M., & Ushijima, T. (1978). Laryngeal control in French stop production: A fiberscopic, acoustic and electromyographic study. *Folia Phoniatrica*, *30*, 175-198.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*, 129-159.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, *10*, 177-192.

Frøkjaer-Jensen, B., Ludvigsen, C., & Rischel, J. (1973). A glottographic study of some Danish consonants. *Annual Report* (Institute of Phonetics, University of Copenhagen), *7*, 269-295.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, *55*, 37-58.

Jeel, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. *Annual Report* (Institute of Phonetics, University of Copenhagen), *9*, 191-211.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics*, *2*, 161-180.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*, 286-319.

Kelso, J. A. S., & Tuller, B. (1985). Intrinsic time in speech production: Theory, methodology, and preliminary observations. *Haskins Laboratories Status Report on Speech Research*, *SR-81*, 23-39.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics, University of Stockholm*, *21*.

Lindqvist, J. (1972). Laryngeal articulation studied on Swedish subjects. *Quarterly Progress and Status Report* (Speech Transmission Laboratory, University of Stockholm), *2-3*, 10-27.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

Lisker, L., Abramson, A. S., Cooper, F. S., & Schvey, M. H. (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America*, *45*, 1544-1546.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America*, *71*, 173-178.

Petersen, N. R. (1983). The effect of consonant type on fundamental frequency and larynx height in Danish. *Annual Report* (Institute of Phonetics, University of Copenhagen), *17*, 55-86.

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction II. *Language and Speech*, *12*, 137-155.

Slis, I. H., & Damste, P. H. (1967). Transillumination of the glottis during voiced and voiceless consonants. *IPO Annual Progress Report*, *2*, 103-113.

MAINSTREAMING MOVEMENT SCIENCE*

J. A. S. Kelso†

The target article by Berkenblit, Fel'dman, and Fukson (in press) is a fine synthesis of a program of research that attacks many of the important issues facing movement science. Our view is that some of these issues--local though they may seem to the study of movement--can be usefully viewed within a larger scientific context, particularly, recent developments in nonlinear dynamical systems and theories of cooperative phenomena in physical, chemical, and biological systems. Thus, the multidegree of freedom movements of animals and people--whose principles continue to elude us--may be couched within, or be extensions of, the laws underlying order and regularity in other natural systems. Thereby, a minimum set of principles may emerge for understanding utterly diverse phenomena (Maxwell, 1877).

As others have long realized (e.g., Boylls, 1975; Greene, 1971, 1982; Turvey, 1977), the field of control and coordination of movement has a rich heritage, stemming from Bernstein's influential theories and empirical research. In a sense, in the West we are only beginning to appreciate Bernstein's legacy, an appreciation not only forced on us by emerging data on multidegree-of-freedom activities, but also as we get a better feel for the deep problems of biological motion. For example, recent work on human posture has demonstrated that rapid and flexible reactions occur in remote muscles when those activities are necessary to preserve *function* (e.g., stable posture when holding a cup of tea). Claims, however, that such effects "constitute a distinct and apparently *new*, class of motor reaction" (Marsden, Merton, & Morton, 1983, p. 645, emphasis ours) are myopic in light of this and previous Russian work, and may simply reflect a Western bias (see e.g., Gelfand, Gurfinkel, Tsetlin, & Shik, 1971). The old aphorism that one who is ignorant of history is destined to relive it, applies also, it seems, to insular attitudes in science.

It is interesting to note that Marsden et al.'s research on posture led them, by their own admission, to abandon an earlier influential servo-theory

---

*The author was invited by the Editor of *The Behavioral and Brain Sciences* to publish this article as a *Continuing Commentary* on the Berkenblit, Fel'dman, and Fukson (in press) article, but declined. The present paper was considered too long for inclusion in the original treatment, but was nevertheless forwarded to the authors, who have considered it in their *Response to Commentators* (Fel'dman, personal communication).

†Also Center for Complex Systems and Department of Psychology, Florida Atlantic University, Boca Raton.

Acknowledgment. Much of the research discussed herein was supported in part by NIH Grant NS-13617, BRS Grant RR-05596, and Contract No. N0014-83-C-0083 from the U. S. Office of Naval Research.

[HASKINS LABORATORIES: Status Report on Speech Research SR-85 (1986)]

of stretch reflex function. Yet Berkenblit et al. hold tightly to the concept of reflex as the basis of volitional action, even though the frog's wiping behavior is far from the "machine-like fatality" envisaged by Sherrington or the "machine-like, inevitable reactions" of Pavlov (see Fearing, 1930/1970). In the first part of this comment, we advocate a requiem for the reflex, Sherrington's (1906) "likely, if not probable fiction" and "purely abstract conception." We take, along with much other evidence, the adaptive, context-sensitive and functionally-specific motor behavior of the spinal frog--beautifully shown in the experiments of Berkenblit et al.--as contributing for the reflex's death-knell. In a manner consistent with Bernstein, we claim that the functional units of action are not anything like reflexes: reflexes may be elemental, but they are not fundamental in the sense of affording an understanding of coherent action (for a discussion of the elemental-fundamental distinction in modern particle physics, see Buckley & Peat, 1979). The reflex is a vestige of Descartes and Newton, of a machine-view of animal action, and in our view it is time to discard it. The same could be said of explanations whose ontology rests on the formal machine concept, that is, the motor program and its neurally-based counterpart, the central pattern generator (CPG). But that issue has been addressed elsewhere as Berkenblit et al. note (Footnote 1, see also Kelso, 1981, in press; Selverston, 1980, and commentaries).

The second part of this commentary addresses two central issues lucidly demonstrated and discussed by Berkenblit et al.: (1) the capability of a tremendously complex system possessing a huge number of degrees of freedom to "simulate" a simple, knowable system like a mass-spring; and (2) relatedly, the system's capability to achieve the same macroscopic product (e.g., wiping) with a variety of different effectors, in the face of perturbations and changes in initial conditions, and through a (potentially infinite) number of trajectories. As Berkenblit et al. note, this "constancy" has parallels in perception and even in morphogenesis. The reproducibility of functional behavior in spite of much variability in the "reflex" itself and in the components that contribute to it is indicative of what the biologist would call structural stability, that is, a pronounced invariance in form and function against spatial or temporal deformations (e.g., Thom, 1975; Thompson, 1917; Weiss, 1969). These *facts* of action and perception (not principles, mark you--phenomena to be understood) can be brought under a common dynamical framework, although here we can only hint at its general features. To some extent, this involves linking the work of Berkenblit et al., with that of their colleagues who study regular and stochastic motion in simple and multidegree of freedom dissipative systems (e.g., Andronov & Chaikin, 1949; Arnol'd, 1978)--a field in which there is currently tremendous interest (e.g., Feigenbaum, 1980; Grassberger & Procaccia, 1983; Haken, 1975, 1983).

Requiem for the Reflex?

The idea that voluntary movement is constructed from reflexes (innate patterns) and ultimately effected by reflex parameterization is not new (cf. Fearing, 1930/1970), although the particular mechanism envisaged by Berkenblit et al. may be. Much confusion has arisen in physiology and psychology over the usage, meaning, and assumptions underlying the concept of reflex. It would not be too hard to document, in the fifty-five years following Fearing's brilliant historical and critical treatise on the reflex, the same conceptual pitfalls that he detailed.

A chief source of confusion rests on the assumption that simplicity of anatomy dictates simplicity of function, and vice-versa. The reflex arc is something with which we are all familiar, and the step toward understanding stimulus-response behavior must have seemed a natural one. In fairness, Sherrington (1906) was wary of such interpretative ease; near the end of his Silliman lectures, he stressed the importance of understanding the interaction of reflexes and volitional control. In our view, Sherrington was in a bind. On the one hand, he could map neurophysiologically the "reflex machinery," the wiring diagram, in certain simple cases (e.g., spinal preparations) and thus relate anatomy to function (e.g., the scratch reflex, the stepping reflex). On the other hand, as a self-professed Darwinian, he recognized that "the difficulty in assigning purpose to a particular reflex is hazardous and inversely proportional to the field covered by the reflex effect" (Sherrington, 1906, p. 239). In our view, the difficulty lay in Sherrington's belief that reflexes, by definition, were hard-wired entities.

Bernstein (1928/1967) took a very different tack from Sherrington. For Bernstein, movement was hypothesized to be "a living morphological object," not "chains of details but structures which are differentiated into details" (p. 67). The identity of movement with emerging form meant that changes in one single detail of a movement could lead to "a whole series of others which are sometimes very far removed from the former both in space and time" (p. 69). For Bernstein, a perturbation to the "motor field" was felt by the field as a whole in such a way as to preserve integrity of system function. It was the form or topology of action that was preserved. This is the essence of the coordinative structure construct (Berkenblit et al., in press, Section 3), by definition, an ensemble of neuromuscular components temporarily assembled as a functionally-specific unit. The remarkable adaptability and variability in the spinal frog's wiping behavior is characteristic of a coordinative structure, not a reflex--at least by any conventional definition.

A coordinative structure organization--as seen in the spinal frog--is apparent in many different activities attesting further to Berkenblit et al's intuition that nature operates with ancient themes. But in our view, it is not so much that higher levels exploit innate patterns as it is that coordinative structures are evident at every level of motor system description and across phyletic strata. This is because functions, not reflexes, are evolutionary primitives. For example, in the case of speech, a so-called "higher level" activity, an unexpected perturbation to the jaw during upward motion for final /b/ closure in the utterance /bæb/ reveals near-immediate changes in upper and lower lip muscles and movements (15-30 ms), but no changes in tongue muscle activity. The same perturbation applied during the utterance /bæz/ evokes rapid and increased tongue muscle activity for /z/ frication, but no active lip movement (Kelso, Tuller, & Fowler, 1982; Kelso, V. Bateson, & Fowler, 1984). Note that the form of interarticulator coordination is neither random nor hard-wired, but unique and specific to the phoneme produced. That a challenge to one member of a group of potentially independent articulators is met--on the very first perturbation experience--by remotely (but not, note, mechanically) linked members of the group, provides strong support for coordinative structures as the meaningful units of behavioral action, regardless of anatomical "level." Though such adaptive behavior could, because of its speed, be described as reflexive, its mutability speaks against any kind of reflex organization.

To recognize the coordinative structure as Greene's (1971) "significant informational unit" is not merely a plea for a change in terminology. It is to underscore the "soft," flexible nature of a unit of action, and to take us away from the hard-wired language of reflexes and CPGs or the hard-algorithmed language of computers (formal machines), which are the source of the motor program/CPG idea. In place of such machine metaphors, the coordinative structure construct emphasizes the analytic tools of qualitative (nonlinear) dynamics (e.g., Abraham & Shaw, 1982; Arnol'd, 1978) and the physical principles of cooperative phenomena in nonequilibrium, open systems (e.g., Haken, 1975). It is this "equipment" that may, on the one hand, provide a principled account of the phenomena discussed by Berkenblit et al. (in press) and, on the other, bring the study of biological motion into the mainstream of theoretical science.

Constancies, Motor Equivalence and Attractors

Berkenblit et al. (in press) ask: How do different movement trajectories, with different effectors and in the face of changing contextual conditions, manage to accomplish the same goal? Similarly, for the case of perceptual constancies, one can inquire: How do different retinal images yield the same percept? Note that in each case the number of microscopic degrees of freedom is enormous (e.g., the neurons, neuronal connections, muscle fibers involved in lifting a finger or the light rays to the eye, the retinal mosaic, and neural processing structures involved in perceiving an object). Yet somehow, this high dimensionality gets "compressed" into a lower dimensional subspace. How such compression is realized is the challenge faced, not only by students of action and perception, but in other realms of science as well. For example, chemistry asks how low-dimensional behavior, such as periodicity, arises in the Belousov-Zhabatinskii reaction even when thirty or more chemical species are present in the reaction vessel (e.g., Shaw, 1981). In the case of movement control and perception, a key may lie in the identity between the flow of a dynamical system (as reflecting, say, the self-equilibrating characteristic of a complex, multidegree of freedom motor system) and the flowing optic array described originally by Gibson (1950). In the former case, the flow is represented in the qualitative shapes or forms of motion observed in the system's phase portrait, that is, the totality of all possible phase plane trajectories generated by a particular dynamical system under a given parameterization. In the latter case, the visual flow is equivalent to optical structure (defined in terms of optical motion vectors rather than Euclidean images, see e.g., Johansson, 1977) that is lawfully generated by the environmental layout of surfaces and by the movements of animals (see e.g., Gibson, 1950, Chapter 7). In each case the relevant parameters are found to be macroscopic and low-dimensional.

Tasks like wiping off a noxious stimulus or reaching for a cup yield patterned forms of motion characteristic of point attractor dynamics, a generic category that denotes the fact that all trajectories on the phase portrait flow to an asymptotic equilibrium state (a basin of attraction). It is important to realize that multidegree-of-freedom systems whose trajectories converge to a stable position can also be described in the low-dimensional language of point attractor dynamics. In the context of movement, this is because the system is dissipative, that is, there is a contraction (not a conservation as in Hamiltonian systems) of phase space volume onto a surface of lower dimensionality than the original space. Other kinds of attractors corresponding to stable, steady-state motions in N-dimensional systems are periodic attractors or limit cycles, which are capable of characterizing rhythmical tasks like chewing, locomotion and perhaps speaking (e.g., Kelso, V.-Bateson, Saltzman, & Kay, 1985). Moreover, given the presence of chaos even in simple deterministic dynamical systems (e.g., Feigenbaum, 1980), chaotic attractors in movement are not unlikely. Here we see the beginning of a way to conceptualize and model how an extremely complex system becomes controllable as low-dimensional dynamics.

This is obviously only a small part of a big story. Berkenblit et al. (in press) refer in several places to critical behavior (e.g., the critical hip phase angle for initiating the locomotory swing phase) and bifurcations (e.g., in terms of switching among trajectory subcomponents, abrupt modifications of movement pattern). As they note, although such phenomena are well known (if under-recognized) in movement, their lawful basis is not understood. Certain theoretical programs that deal explicitly with pattern formation and change (e.g., Haken, 1975; Nicolis & Prigogine, 1977) suggest a basis for understanding these and other phenomena. In synergetics, for example, near regions of instability (i.e., before qualitative shifts in pattern occur) the system's behavior can be completely specified by one or a few order parameters (Haken, 1975, 1983). Such order parameters are created by the cooperation among the individual components of a complex system, and they in turn govern the behavior of these components. They therefore afford, in principle, a linkage between macro- and microlevels of description. Using concepts of synergetics and nonlinear oscillator theory, Haken, Kelso, and Bunz (1985) have offered an explicit theoretical model of phase transitions in bimanual activity (Kelso, 1984) that should have general applicability to the kinds of critical phenomena and bifurcations discussed by Berkenblit et al. (in press). The theoretical strategy employed by Haken et al. may be worth noting. First, they specify the layout of attractor states characterizing the stable bimanual modes and show how, under the influence of continuous scaling on a control parameter, the layout changes--at a critical value--from one attractor to another. Then they derive this scenario and other features of the data (Kelso, 1984) from the equations of motion of each hand and a nonlinear coupling between them. Recently, Kelso and Scholz (1985) have verified several novel predictions of an extended version of the model (Schöner, Haken, & Kelso, 1986), including the existence of critical slowing down in order parameter behavior as the transition is approached, and enhanced fluctuations in order parameter behavior near the bifurcation region. Such predictions and results would hardly be expected from conventional motor program/CPG accounts of "switching" behavior, for example, gait changes (cf. Grillner, 1982, p. 224; Schmidt, 1982, p. 316).

The present framework may apply not just to biological motion, *per se*, but to the perception-action system as a total unit. Elaboration of Gibson's work on visual flow fields for example (see e.g., Lee, 1980; Lee & Reddish, 1981) reveals how the rate of dilation, $\tau(t)$ of a bounded region of optical structure specifies the time at which a moving object will contact a surface. (Note: the ratio of retinal expansion velocity and retinal size is equivalent to the inverse of τ). Flies, for example, have been demonstrated to begin to decelerate prior to surface contact at a critical value of the inverse of τ (Wagner, 1982). Thus, not only does the τ parameter and its rate of change, $\dot{\tau}$, provide continuous information for modulating ongoing activity, but at certain critical points, the system exhibits bifurcations to adaptive modes of behavior as well. In this view, then, information for the perception-action system is specified in the morphology of the flow field (Gibson, 1950, Ch. 7).

The flow field geometry is defined in terms of flow vectors to and from a focus of expansion, which can be conceived as basically an attractor or repellor. As the facts of motor equivalence/equifinality tell us, attractors and their layout must be defined in terms of their significance or meaningfulness for the perception-action system, (i.e., in Berkenblit et al., in press, the "reflex" is variable, but the goal achievement is not). Thus, a further consequence of the present framework, to which we can only allude here, is a dynamic information theory (see e.g., Haken, 1984)--one in which information is not viewed in the classical Shannonian sense, as a measure for scarcity of a message or ignorance regarding systemic states (i.e., as receiver-independent), but rather as carrying its own semantic content for the receiver.

In conclusion, there is reason to suppose that an understanding of the multidegree of freedom activities of animals and people falls squarely on the shoulders of an emerging theory of cooperativity and pattern formation in open, complex systems. If so, we conclude where we began, namely that many of the phenomena beautifully treated by Berkenblit et al. (in press) may not be "special" to movement science, and thus may not require "special" concepts beyond those developed from first principles. The view that we are pursuing is that biological motion is an important test field for the essential elaboration of these basically physical (but, mark, non-Newtonian and nonmechanical) ideas.

References

Abraham, R. H. & Shaw, C. D. (1982). Dynamics--The geometry of behavior. Santa Cruz, CA: Aerial Press.

Andronov, A., & Chaiken, C. E. (1949). Theory of oscillations. Princeton, NJ: Princeton University Press.

Arnol'd, V. I. (1978). Mathematical methods of classical mechanics. New York: Springer Verlag.

Berkenblit, M. B., Fel'dman, A. G., & Fukson, O. I. (in press). Adaptability of innate motor patterns and motor control mechanisms. The Behavioral & Brain Sciences.

Bernstein, N. A. (1967). The coordination and regulation of movements. London: Pergamon Press.

Boylls, C. C. (1975). A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat (COINS Technical Report 76-1). Amherst, MA: University of Massachusetts, Department of Computer and Information Science.

Buckley, P. F., & Peat, F. D. (1979). A question of physics. Toronto: University of Toronto Press.

Fearing, F. (1930/1970). Reflex action: A study in the history of physiological psychology. Cambridge, MA: MIT Press.

Feigenbaum, M. J. (1980). Universal behavior in nonlinear systems. Los Alamos Science, 1, 4-27.

Gelfand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. (1971). Some problems in the analysis of movements. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge: MIT Press.

Gibson, J. J. (1950). The perception of the visual world. Boston: Houghton-Mifflin.

Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. Physica, 9D, 189-208.

Greene, P. H. (1971). Introduction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, MA: MIT Press.

Greene, P. H. (1982). Why is it easy to control your arms? Journal of Motor Behavior, 4, 260-286.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control. Oxford: Pergamon Press.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. Review of Modern Physics, 47, 67-121.

Haken, H. (1983). Advanced synergetics: Instability hierarchies of self-organizing systems and devices. Heidelberg: Spring-Verlag.

Haken, H. (1984). Towards a dynamic information theory. In I. Lamprecht & A. I. Zotin (Eds.), Thermodynamics and regulation of biological processes. Berlin: de Gruyter.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51, 347-356.

Johansson, G. (1977). Spatial constancy and motion in visual perception. In W. Epstein (Ed.), Stability and constancy in visual perception. New York: Wiley.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative, and Comparative, 15, R1000-R1004.

Kelso, J. A. S. (in press). Pattern formation in multidegree of freedom speech and limb movements. In H. Heuer, C. Fromm, C. Brunia, J. A. S. Kelso, & R. A. Schmidt (Eds.), Generation and modulation of action patterns. Experimental Brain Research Supplement, 15.

Kelso, J. A. S., & Scholz, J. P. (1985). Cooperative phenomena in biological motion. In H. Haken (Ed.), Complex systems: Operational approaches in neurobiology, physics and computers (pp. 124-149). Heidelberg: Springer-Verlag.

Kelso, J. A. S., Tuller, B., & Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. Journal of the Acoustical Society of America, 72, S103.

Kelso, J. A. S., Tuller, B., V.-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. *Nature*, *293*, 293-294.

Marsden, C. D., Merton, P. A., & Morton, H. B. (1983). Rapid postural reactions to mechanical displacement of the hand in man. In J. E. Desmedt (Ed.), *Motor control mechanisms in health and disease* (pp. 645-659). New York: Raven Press.

Maxwell, J. C. (1877). *Matter and motion*. New York: Dover Press.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley-Interscience.

Schmidt, R. A. (1982). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics.

Schöner, G., Haken, H., & Kelso, J. A. S. (1986). A stochastic theory of phase transitions in human hand movement. *Biological Cybernetics*.

Selverston, A. I. (1980). Are central pattern generators understandable? *The Behavioral and Brain Sciences*, *3*, 535-571.

Shaw, R. (1981). Modeling chaotic systems. In H. Haken (Ed.), *Order and chaos in nature*. Heidelberg: Springer Verlag.

Sherrington, C. S. (1906). *The integrative action of the nervous system*. London: Constable.

Thom, R. (1975). In D. H. Fowler (Trans.), *Structural stability and morphogenesis*. Reading, MA: Benjamin, Inc.

Thompson, D. W. (1917). *On growth and form*. London: Cambridge University Press.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, NJ: Erlbaum.

Wagner, H. (1982). Flow-field variables trigger landing in flies. *Nature*, *297*, 147-148.

Weiss, P. A. (1969). The living system: Determinism stratified. In A. Koestler & J. R. Smythies (Eds.), *Beyond reductionism* (pp. 3-42). Boston: Beacon.

PUBLICATIONS
APPENDIX

PUBLICATIONS

Abramson, A. S. (in press). The Thai tonal space. *Proceedings of the 18th International Conference on Sino-Tibetan Languages and Linguistics*, Bangkok, August 27-29, 1985.

Alfonso, P. J., & Seider, R. A. (1986). Laryngeal and respiratory physiological characteristics of inaudible and audible dysfluent production. In S. Hibi, D. Bless, & M. Hirano (Eds.), *Proceedings of the International Congress on Voice* (pp. 13-22). Kurume: Kurume University.

Beddor, P., S., Krakow, R. A., & Goldstein, L. M. (in press). Perceptual contraints and phonological change: A study of nasal vowel height. *Phonology Yearbook*, *Vol. 3*.

Bingham, Geoffrey P. (in press). Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*.

Browman, C. P., & Goldstein, L. M. (in press). Towards an articulatory phonology. *Phonology Yearbook* (Vol. 3, 1986).

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (1986). P-centers are unaffected by phonetic categorization. *Perception & Psychophysics*, *39*, 187-196.

Feldman, L. B. (in press). Phonological and morphological analysis by skilled readers of Serbo-Croatian. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language Perception and Production*. London: Academic Press.

Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3-28.

Fowler, C. (1986). Reply to commentators. *Journal of Phonetics*, *14*, 149-171.

Goldstein, L., & Browman, C. P. (in press). Representation of voicing contrasts using articulatory gestrues. *Journal of Phonetics*.

Hanson, V. L. (1986). Access to spoken language and the acquisition of orthographic structure: Evidence from deaf readers. *Quarterly Journal of Experimental Psychology*, *38A*, 193-212.

Kelso, J. A.S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, *14*, 29-59.

Mann, V. A. (in press). Phonological awareness: The role of reading experience. *Cognition*.

Nittrouer, S., & Hochberg, I. (1985). Speech instruction for deaf children: A communication-based approach. *American Annals of the Deaf*, *130*, 491-495.

Nittrouer, S., & Studdert-Kennedy, M. (in press). The stop-glide distinction: Acoustic analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/. *Journal of Acoustical Society of America*.

Nittrouer, S., & Studdert-Kennedy, M. (in press). The role of coarticulation in the perception of speech by young children. *Journal of Speech and Hearing Research*.

Repp, B. H. (1985). Can linguistic boundaries change the effectiveness of silence as a phonetic cue? *Journal of Phonetics*, *13*, 421-433.

Repp, B. H. (1986). Some observations on the development of anticipatory coarticulation. *Journal of the Acoustical Society of America*, *72*, 1616-1619.

Repp, B. H. (1986). Perception of the [m]-[n] distinction in CV syllables. Journal of the Acoustical Society of America, 79, 1387-1999.

Repp, B. H. (in press). The role of psychophysics in understanding speech perception. In M. E. H. Schouten (Ed.), Proceedings of the NATO Advanced Research Workshop, The Psychophysics of Speech Perception, Utrecht, The Netherlands.

Richards, J. T. & Hanson, V. L. (1985). Visual and production similarity of the handshapes of the American manual alphabet. Perception & Psychophysics, 38, 311-319.

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. Journal of Memory and Language, 25, 123-135.

Studdert-Kennedy, M. (1986). Two cheers for direct realism. Journal of Phonetics, 14, 99-104.

APPENDIX

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834
SR-71/72	July - December 1982	AD A124596	ED 225-212
SR-73	January - March 1983	AD A129713	ED 229-816
SR-74/75	April - September 1983	AD A136416	ED 236-753
SR-76	October - December 1983	AD A140176	ED 241-973
SR-77/78	January - June 1984	AD A145585	ED 247-626
SR-79/80	July - December 1984	AD A151035	ED 252-907
SR-81	January - March 1985	AD A156294	ED 257-159
SR-82/83	April - September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
SR-85, 1986

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Haskins Laboratories Status Report
on Speech Research

5. TYPE OF REPORT & PERIOD COVERED
Continuation Report
Jan. 1-March 31, 1986

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Staff of Haskins Laboratories:
Alvin M. Liberman, President

8. CONTRACT OR GRANT NUMBER(s)

HD-01994	NS13870
N01-HD-5-2910	NS13617
RR-05596	NS18010

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Haskins Laboratories
270 Crown Street
New Haven, CT 06511-6695

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
National Institutes of Health
National Science Foundation
Office of Naval Research

12. REPORT DATE
March, 1986

13. NUMBER OF PAGES
280

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
As above

15. SECURITY CLASS. (of this Report)
Unclassified

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
UNLIMITED: Contains no information not freely available to the general public. It is distributed primarily for library use.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
As above

18. SUPPLEMENTARY NOTES
N/A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Speech Perception:
constraints, phonological change, nasals, vowel height
reading ability, p-centers, categories, events,
direct-realism, short-term memory, acoustic analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report (1 January-31 March) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

-Phonological awareness: The role of reading experience
-An investigation of speech perception abilities in children who differ in reading skill
-Phonological and morphological analysis by skilled readers of Serbo-Croatian
-Visual and production similarity of the handshapes of the American Manual Alphabet
-Morphophonology and lexical organization in deaf readers

DD FORM 1473 1 Jan 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

8. Contract or Grant Numbers (Continued)

BNS-8111470
BNS-8520709
N00014-83-K-0083

19. Key Words (Continued)

Speech Articulation:
articulatory phonology, voicing contrast, gestures, velotrace, velum, apparatus, dynamical perspective, acoustic analysis, coarticulation, Catalan, Spanish, vowels, consonants

Motor Control:
movement-science

Reading:
Serbo-Croatian, phonology, analysis, words, print, deaf signers, sign coding, morphophonology, lexical organization, phonological awareness, sign similarity, handshapes, manual alphabet

20. Abstract (Continued)

-Short-term memory for printed English words by congenitally deaf signers: Evidence of sign-based coding reconsidered
-Perceptual constraints and phonological change: A study of nasal vowel height
-The Thai tonal space
-P-centers are unaffected by phonetic categorization
-Two cheers for direct realism
-An event approach to the study of speech perception from a direct-realist perspective
-The dynamical perspective on speech production: Data and theory
-The Velotrace: A device for monitoring velar position
-Towards an articulatory phonology
-Representation of voicing contrasts using articulatory gestures
-Mainstreaming movement science